Remembering faces and scenes: the mixed-category advantage in visual working memory

Yuhong V. Jiang 1,2
Roger W. Remington 1,2
Anthony Asaad 1
Hyejin J. Lee 1
Taylor C. Mikkalson
1. Department of Psychology, University of Minnesota
2. Center for Cognitive Sciences, University of Minnesota

Abstract

We examined the mixed-category memory advantage for faces and scenes to determine how domain-specific cortical resources constrain visual working memory. Consistent with previous findings, visual working memory for a display of two faces and two scenes was better than that for a display of four faces or four scenes. This pattern was unaffected by manipulations of encoding duration. However, the mixed-category advantage was carried solely by faces: memory for scenes was not better when scenes were encoded with faces rather than with other scenes. The asymmetry between faces and scenes was found when items were presented simultaneously or sequentially, centrally or peripherally, and when scenes were drawn from a narrow category. A further experiment showed a mixed-category advantage in memory for faces and bodies, but not in memory for scenes and objects. The results suggest that unique category-specific interactions contribute significantly to the mixed-category advantage in visual working memory.

Keywords: visual working memory; domain specificity

Introduction

Human cognition depends on both specialized and distributed brain processing. The degree of specialization varies, with sensorimotor processing relying more heavily on specialized brain processes than do executive control and other central mechanisms (Gazzaniga, Ivry, & Mangun, 2013). When multiple tasks or stimuli rely on different cortical resources, they tend to produce less interference compared with similar tasks or stimuli. Thus, for example, whereas two visual tasks interfere with each other, visual and auditory tasks may proceed in parallel (Treisman & Davies, 1973). These findings have led to the proposal that shared cortical resources are a primary driver of competition (Just & Varma, 2007; Wickens, 2002).

One line of evidence for the role of shared cortical resources comes from research on visual working memory. Working memory is frequently considered one component of executive function (Miyake et al., 2000; Shipstead, Harrison, & Engle, 2015). Secondary tasks interfere with visual working memory even when they involve auditory processing (Makovski, Shim, & Jiang, 2006). Increasing working memory load, be it color, shape, or motion, produces greater activity in posterior parietal and frontal regions (Culham, Cavanagh, & Kanwisher, 2001; Curtis & D’Esposito, 2003; Song & Jiang, 2006; Todd & Marois, 2004). This reliance on central mechanisms implies that competition for visual working memory should be widespread, even for stimuli initially encoded by distinct brain regions. However, sensory and perceptual regions of the brain actively participate in working memory tasks (for a review, see Eimer, 2015). When participants hold an oriented grating in working memory, V1 neurons retain information about the memorized orientation (Bettencourt & Xu, 2015; Harrison & Tong, 2009). More broadly, the content of visual working memory biases activity in ventral visual areas (Ranganath, DeGutis, & D’Esposito, 2004). After viewing a stream of faces and scenes, participants show greater activity in the fusiform face area when they hold faces in working
memory and greater activity in the parahippocampal place area when they hold scenes in working memory (Ranganath et al., 2004). The reliance on stimulus-specific regions implies that cortical resources should mediate competition for visual working memory: stimuli encoded by more distinct cortical regions should yield less competition.

One recent study used functional Magnetic Resonance Imaging (fMRI) to link enhanced memory performance for mixed-category stimuli with increased cortical resources in the ventral visual stream (Cohen, Konkle, Rhee, Nakayama, & Alvarez, 2014). Cohen et al. (2014) measured the degree of cortical overlap in the ventral occipito-temporal regions for the perception of various categories, including faces, scenes, objects, and bodies. Some categories, such as objects and scenes, exhibit substantial cortical overlap, whereas other categories, such as faces and scenes, rely on more distinct cortical regions. The study also measured the memory advantage for remembering mixed-category stimuli relative to single-category stimuli. Cortical overlap in the perception of two categories correlates significantly with the size of mixed-category memory advantage. For example, the mixed-category memory advantage is greater for faces and scenes than for objects and scenes. These findings provide compelling evidence for the idea that an increase in domain-specific cortical resources facilitates visual working memory (the cortical resource theory; Cohen et al., 2014).

The data reported above, however, do not specify the locus of the mixed-category advantage or its generality. To retain information in working memory, one must first perceive the visual input for memory encoding. Do increased cortical resources facilitate perceptual processing of mixed-category stimuli, or does it enhance other components of memory, such as storage and retrieval? Neurophysiological studies show that unlike prefrontal neurons, neurons in the inferior temporal lobe of monkeys do not maintain sustained activity during the memory retention interval (Miller, Erickson, & Desimone, 1996). This finding suggests that inferior temporal regions may be involved primarily in perceiving rather than retaining stimuli (see further discussion about this in human neuroimaging, Bettencourt & Xu, 2015). In Cohen et al. (2014)’s study, participants must encode four complex stimuli presented for a total of 800ms. This provided just 200ms per item of encoding time, shorter than what may be adequate for encoding complex stimuli into working memory (Brady, Konkle, Oliva, & Alvarez, 2009; Eng, Chen, & Jiang, 2005; Liu & Jiang, 2005). If increased cortical resources facilitate primarily perception under brief viewing conditions, then the mixed-category memory advantage should be strongest when encoding duration is severely limited. In contrast, if increased cortical resources facilitate other memory components (such as retaining multiple stimuli in working memory), then the mixed-category advantage should manifest even when participants have ample time to perceive the stimuli.

A second unanswered question regarding the cortical resource theory is its generality. Cortical resource theory, thus far, makes no distinction between category types, implying that increasing cortical resources should facilitate visual working memory for all categories. Consider faces and scenes. If working memory for faces is primarily limited by face-specific processing regions, then competition for the face-specific resource should be stronger when the display contains four faces rather than two faces and two scenes. Similarly, if working memory for scenes is primarily limited by scene-specific processing regions, then competition for the scene-specific resource should be stronger when the display contains four scenes rather than two faces and two scenes. Therefore, if the cortical resource theory is broadly applicable to all categories of stimuli, then the mixed-category memory advantage should be a general phenomenon. If, on the other hand, the mixed-category advantage reflects within-region interactions, then it may be observed for some categories but not others. Such a finding would suggest that domain-specific cortical interactions, rather than the total pool of resources, limit working memory. Cohen et al. (2014) reported just the average memory performance across categories, making it impossible to determine whether the mixed-category advantage applies to each category.

An increase in cortical resources alone might not adequately capture how mixed-category stimuli are represented in visual working memory. First, if visual working memory is subject to general central resource limitations, then how people remember mixed-category stimuli will depend
on how they allocate central resources. A preference for one category over another should yield a mixed-category advantage for the preferred category and a mixed-category disadvantage for the non-preferred category. Thus, mixing categories can induce a tradeoff between two categories. Second, competition for within-category resources is affected not only by memory load of that category, but also by unique ensemble processing of the specific category. Consider faces. All faces share the same basic features (e.g., eyes, mouth) and first-order configuration (e.g., two eyes above the nose and mouth). This similarity is associated with an automatic extraction of an “average” face from a set, at the cost of remembering individual face identities (de Fockert & Wolfenstein, 2009). Although people readily extract the gist and “spatial envelope” of scenes (Oliva & Torralba, 2006), unlike faces, scenes do not share common features and configurations and therefore are not susceptible to the same type of averaging as faces. Mixing faces and scenes can facilitate face memory, not simply because of increased cortical resources, but because it reduces face-specific interference. Thus, unique interference among some stimuli may yield a mixed-category memory advantage for just these stimuli, without a corresponding advantage for other stimuli.

In sum, a consideration of just cortical overlap between two categories would predict that the mixed-category effect is a general phenomenon. In contrast, asymmetric central resource allocation would yield a memory tradeoff between two categories. Finally, unique category-specific interference would yield a strong mixed-category advantage for some categories but not for others.

We report four experiments that addressed the origin and generality of the mixed-category advantage in visual working memory. We tested faces and scenes in most experiments because data from these categories provided the strongest evidence for the cortical resource theory (Cohen et al., 2014). To examine whether the mixed-category advantage has primarily a perceptual locus, we varied encoding time: 800ms or self-paced. To test the generality of the mixed-category advantage, we measured memory accuracy for each category (e.g., faces or scenes) as a function of whether it was encoded with the other category or with the same category. This allowed us to test not just whether overall memory performance was better for mixed-category than single-category stimuli, but more specifically whether the mixed-category advantage applied to both categories. Varying presentation format (simultaneous or sequential), image diversity (random scenes or sub-category scenes), and stimulus location (central or peripheral) helped us rule out alternative interpretations. A final experiment examined the specificity of the mixed-category advantage by testing two additional categories: bodies and objects.

**Experiment 1**

We aimed to replicate Cohen et al. (2014)’s mixed-category advantage for faces and scenes using the same stimuli and presentation duration as what they used. In addition, to examine whether the effect was restricted to brief presentation duration, we tested another condition in which participants were allowed to view the encoding display for as long as they wanted. If the mixed-category advantage occurred primarily because brief presentation duration impairs the perception of single-category stimuli, then this effect should be abolished in the self-paced encoding condition.

To verify that encoding duration could limit working memory for faces and scenes, in a pilot study we systematically manipulated the encoding duration for single-category stimuli. Stimuli were drawn from Cohen et al.’s (2014) set. Participants (N=20) encoded either 4 faces or 4 scenes to memory; the encoding duration was 200, 400, 800, 1600, or 3200ms. We found that within the range of duration tested, every doubling of encoding duration yielded a performance improvement of about 4%. These data showed that memory errors at short durations could be due to inadequate encoding of the stimuli. They raised the possibility that the mixed-category advantage could originate from encoding difficulties when participants had just 800ms to encode the displays. Experiment 1 directly tested this possibility. By measuring face memory and scene memory separately, this experiment also examined the generality of the mixed-category effect.

**Method**
**Participants.** All participants in this study were college students between the age of 18 and 26 years old. They were naïve to the purpose of the study. Each participant completed one experiment. All participants had normal or corrected-to-normal visual acuity and normal color vision.

Twenty participants, including 16 females and 4 males with a mean age of 19.2 years, completed Experiment 1.

**Equipment.** Participants were tested individually in a room with normal interior lighting. They sat at an unrestrained distance of about 40cm away from a computer CRT display. The screen resolution was 1024x768 pixels with a refresh rate of 75Hz. Experiments were coded with MATLAB (www.mathworks.com) and Psychtoolbox (Brainard, 1997; Pelli, 1997).

**Stimuli.** We used Cohen et al. (2014)’s stimuli, including 40 faces and 40 scenes (see the Supporting Information of Cohen et al., 2014). Each image subtended 6°x6° and was in gray scale. The background was gray. All images were presented at full contrast [footnote1].

![Figure 1. A schematic illustration of the working memory task. Participants had 800ms or unlimited time to encode four images to memory. After a 1000ms blank retention interval, they viewed the test display with one image cued by the red frame. Participants judged whether the cued test image was the same or different from the image in that location before.](image)

**Procedure.** Following 20 trials of practice, participants completed the main experiment. The experiment was divided into 10 blocks of 32 trials each. At the beginning of each block participants were informed whether the viewing time would be computer-limited or self-paced. Each trial started with a red fixation point (0.55°x0.55°) for 300ms, followed by the memory display of four images, one in each visual quadrant. The center of the four images formed an imaginary rectangle (size 15.4°x7.5°; e.g., the center coordinate of the upper right image was [7.7°, 3.8°]). In 5 blocks the encoding display terminated after 800ms. In 5 other blocks it stayed until participants pressed the spacebar to terminate encoding. Limited and self-paced encoding blocks alternated and their order was counterbalanced across participants. The encoding display was followed by a blank retention interval of 1000ms. The

---

1 Unlike Cohen et al. (2014), we did not adjust the transparency level of the images for each participant to titrate their memory accuracy for faces and scenes. This is because although making the stimuli highly transparent can lower performance, this manipulation equates encoding difficulty of transparent stimuli with retention difficulty of images that are hard to remember. We therefore used full-contrast stimuli. As reported later, faces and scenes at full contrast produced comparable levels of accuracy, alleviating concerns about intrinsic differences in their memorability.
test display then appeared, containing four images. Three of the items were identical to those in the corresponding locations on the encoding display. The fourth item, enclosed by a red outline rectangle (8° x 8°), was either the same as before (50% of the trials) or a different item of the same category (50% of the trials). The location of the relevant test item was random. Participants were asked to remember the encoding items, and to evaluate whether the test item enclosed by the red rectangle was the same or different from the one in that location before. They pressed “s” or “d” to make a response. We emphasized accuracy but not speed of response. A rising tone followed each correct response and a low buzz followed each incorrect response. The next trial started 1,000ms later. Figure 1 illustrates the trial sequence.

Design. In addition to manipulating the encoding duration (800ms or self-paced), we varied the nature of encoding displays. Within each block of trials, 25% of the trials contained 4 faces on the encoding display, 25% of the trials contained 4 scenes, and the rest of the trials contained 2 faces and 2 scenes presented in random locations. For the mixed-category trials, half of the time the relevant test item was a face, hence probing face memory, and the other half of the time the relevant test item was a scene. Thus, when face memory was tested, the encoding display had either 4 faces (single-category) or 2 faces and 2 scenes (mixed-category). Similarly, when scene memory was tested, the encoding display had either 4 scenes (single-category) or 2 faces and 2 scenes (mixed-category). Scenes and faces presented on each trial were randomly selected from the entire set of 40 scenes and 40 faces.

Results

![Figure 2. Encoding duration in the self-paced blocks. Left: Experiment 1. Right: Experiment 2. Error bars show ±1 S.E. of the mean across participants.](image)

We first examined the amount of time that participants took to encode the displays in the self-paced blocks. Figure 2 (left) plots the encoding duration, separately for different types of stimuli (all faces, mixed-category, or all scenes). Encoding duration differed significantly across stimuli, $F(2, 38) = 4.10, p < .03, \eta^2 = .18$. Participants viewed displays of 4 faces significantly longer than displays of 2 faces and 2 scenes, $t(19) = 2.60, p < .02$, and marginally longer than displays of 4 scenes, $t(19) = 1.78, p = .09$. The latter two conditions did not differ in viewing duration, $t(19) = 0.49, p > .50$. Thus, when allowed to control encoding duration, participants viewed the displays for around 3.5 to 4 seconds, substantially longer than the experimenter-limited time of 800ms. But was the mixed-category advantage specific to the shorter encoding duration?

To examine the locus of the mixed-category advantage, we measured memory accuracy for faces and scenes. An encoding display containing 2 faces and 2 scenes (the mixed-category displays) was classified as a “face” trial if a face was the test item, or a “scene” trial if a scene was the test item. Our analysis therefore separated memory for faces and scenes as a function of whether they were encoded with stimuli from the same category (single-category) or the other category (mixed-category). This analysis was done separately for blocks in which encoding was limited to 800ms and when it was self-paced. Figure 3 plots the data.
We performed a repeated-measures ANOVA using duration (800ms or self-paced), category heterogeneity (mixed or single category), and stimulus material (faces or scenes) as factors. Accuracy was higher in the self-paced blocks than the 800ms-duration blocks, $F(1, 19) = 81.57, p < .001, \eta^2_p = .81$. Replicating Cohen et al. (2014), we found a significant main effect of category heterogeneity, with higher accuracy in the mixed-category than single-category condition, $F(1, 19) = 6.05, p < .03, \eta^2_p = .24$. The mixed-category advantage was found in both durations, resulting in a lack of interaction between duration and category heterogeneity, $F < 1$. However, the mixed-category advantage differed for faces and scenes, yielding a significant interaction between category heterogeneity and stimulus material, $F(1, 19) = 21.12, p < .001, \eta^2_p = .53$. The interaction occurred in both duration conditions, resulting in a lack of three-way interaction, $F(1, 19) = 1.13, p > .30$. The other effects were not significant.

To understand the interaction between category heterogeneity and stimulus materials, we separately analyzed data for faces and scenes. Memory for faces was better when faces were presented with scenes rather than with other faces, $F(1, 19) = 23.99, p < .001, \eta^2_p = .56$, and this effect did not interact with encoding duration, $F(1, 19) = 1.15, p > .25$. Memory for scenes, however, was numerically worse when scenes were presented with faces than with other scenes. This effect was not significant, $F(1, 19) = 1.19, p > .25$, and did not interact with duration, $F < 1$.

The pattern of results reported above held when we used $d’$, rather than percent correct, as the dependent measure of memory performance. This was also the case for subsequent experiments.

**Discussion**

Two main findings emerged from Experiment 1. First, the mixed-category advantage in visual working memory was found for faces but not for scenes. Second, this effect was not restricted to conditions when the encoding duration was short. Although participants preferred to view the encoding display for 3–4 seconds rather than the experimenter-imposed 800ms, the mixed-category advantage was evident in both conditions. The data therefore generalized Cohen et al.’s (2014) findings from short encoding duration to long encoding duration. It suggests that the mixed-category effect is unlikely attributable to impaired encoding of single-category stimuli under brief presentation conditions.

The lack of a mixed-category advantage for scenes suggests that the degree of cortical overlap is not by itself an adequate account of the mixed-category effect. Had the total amount of cortical resource been critical, competition for scene-processing regions should have been reduced when scenes were encoded with faces rather than with other scenes. Yet this did not translate into a memory gain for scenes.

One potential explanation for the difference between faces and scenes lies in stimulus heterogeneity. Though researchers frequently lump all faces into one category and all scenes into another, scenes are arguably more diverse than faces. Few stimuli in the visual environment are faces,
but all images that have some sort of spatial layout can be categorized as scenes. The difference in heterogeneity can be gleaned from the stimuli used in Experiment 1, which included 40 faces and 40 scenes from the set used by Cohen et al. (2014). The 40 scenes contained 10 exemplars from each of four sub-categories: city, ocean, highway, and mountain. A display with four scenes may include images from different sub-categories of scenes. With already distinct scenes, further diversifying them by replacing two of the scenes with faces may not yield a significant increase in cortical resource. In contrast, all faces have the same basic features and first-order configuration. Diversifying the stimuli by replacing faces with scenes is likely to significantly increase cortical resource. By this account, effects of cortical resource on performance come not just from the degree of cortical overlap between two different categories (e.g., faces and scenes), but also from the degree of resource limitation for each category. Category mixing is advantageous primarily for the category whose resource pool is highly limited. As an analogy, one can buy goods with either cash or a credit card. If one has limited cash but unlimited credit, then having both payment methods gives one more purchase power than having just cash, but not more purchase power than having just credit. To find out whether the mixed-category advantage applies to scenes, it would be necessary to use scenes that compete strongly with each other for a limited pool of resources.

Scenes within a sub-category, such as mountains or oceans, are known to share common features, decodable in the ventral visual stream using fMRI (Walther, Caddigan, Fei-Fei, & Beck, 2009). Memory for scenes within a specific category is worse than memory for scenes from diverse categories (Konkle, Brady, Alvarez, & Oliva, 2010). Supporting the idea that visual working memory for sub-category scenes is severely limited, in a pilot study we found that visual working memory for sub-category scenes was poorer than that for random scenes or faces. For this pilot study we measure effects of memory load on three types of stimuli: faces used in Experiment 1, scenes drawn at random from all 40 scenes from Experiment 1, and scenes drawn from the same 40 scenes with the constraint that scenes used in a trial came from a single sub-category (e.g., city, ocean, highway, or mountain). Participants (N=9) encoded 2, 3, or 4 items to memory. Figure 4 displays data from this pilot study.

![Figure 4. Data from a pilot study (N=9) that measured working memory performance as a function of memory load (2, 3, or 4) and stimulus materials (faces, scenes drawn randomly, or scenes drawn from the same sub-category). Error bars show ±1 S.E. of the mean across participants.](image)

An ANOVA showed significant main effects of memory load (p < .001), stimulus materials (p < .002), and their interaction (p < .02). Trend analysis showed that memory for scenes drawn randomly was the best, followed by memory for faces. Memory for scenes drawn from the same sub-category was the worst (p < .002). This finding suggests that scenes drawn from a single sub-category competed strongly for working memory resources. If the size of within-category resource pool is
Visual working memory

important for determining the mixed-category advantage, then this effect should apply as strongly to scenes drawn from a sub-category as to faces. Experiment 2 tested this possibility.

Experiment 2

Instead of randomly selecting scenes from all 40 images, in Experiment 2 we presented scenes from the same sub-category on each trial. Thus, for example, the “all scenes” condition may include 4 ocean scenes, and the “mixed category” condition may include 2 faces and 2 highway scenes. These were the same stimuli as those used in the pilot study, ensuring that scenes on a given display competed strongly for working memory resources. Consequently they may benefit from the increased cortical resource in the mixed-category condition. On the other hand, the greater interference among multiple faces may reflect unique ways in which an ensemble of faces interacts. For example, the type of high-dimensional averaging that occurs automatically to faces (de Fockert & Wolfenstein, 2009) may not occur to scenes, including scenes from the same sub-category. If this is the case, then the mixed-category advantage may remain stronger for faces than for scenes.

Method

Participants. Twenty new participants, including 10 females and 10 males with a mean age of 20 years old, completed Experiment 2.

Design and procedure. This experiment was identical to Experiment 1 except that we constrained the selection of scene images. The set of 40 images used in Experiment 1 (taken from Cohen et al., 2014) comprised of 10 images in each of the four sub-categories: city, ocean, highway, or mountain. Instead of randomly selecting images from all 40 scenes, in Experiment 2 scenes presented on any given trial always came from the same sub-category. The specific sub-category was randomly chosen on each trial and counterbalancing across conditions. The test probe always matched the memory image in its specific sub-category (e.g., an ocean scene would only change into another ocean scene).

Similar to Experiment 1, we manipulated encoding duration in different blocks (800ms or self-paced). Other aspects of the experiment were the same as in Experiment 1.

Results

![Figure 5](image)

Figure 5. Results from Experiment 2. Left: data from blocks in which the encoding display was presented for 800ms. Right: data from blocks in which participants controlled the duration of the encoding display. Error bars show ±1 S.E. of the mean across participants.

In the self-paced blocks participants viewed the encoding displays for about 4s (Figure 2, right). Encoding duration varied across stimuli, $F(2, 38) = 5.17, p < .01, \eta_p^2 = .21$. Participants viewed a display with four faces significantly longer than a display with a mixture of faces and scenes, $t(19) = 2.09, p < .05$, or a display with four scenes, $t(19) = 2.95, p < .008$. The latter two conditions did not differ, $t(19) = 0.76, p > .40$. 


Viewing the memory display longer led to significantly higher accuracy in the self-paced blocks than the limited-duration blocks (Figure 5), $F(1, 19) = 34.10, p < .001$, $\eta^2_p = .64$. In Experiment 2 the scene displays contained scenes from the same sub-category. With this design, memory for scenes was significantly worse than memory for faces, $F(1, 19) = 12.65, p < .002$, $\eta^2_p = .40$. The main effect of mixed versus single category did not reach significance, $F(1, 19) = 2.75, p > .10$. However, just like Experiment 1, category heterogeneity interacted with stimulus material, $F(1, 19) = 16.71, p < .001$, $\eta^2_p = .47$. None of the other effects were significant, all $ps > .10$.

To understand the interaction between category heterogeneity and stimulus material, we separately analyzed data for faces and scenes. Similar to Experiment 1, face memory was better in the mixed-category condition than the single-category condition, $F(1, 19) = 65.56, p < .001$, $\eta^2_p = .78$. This effect did not interact with encoding duration, $F(1, 19) = 1.02, p > .30$. In contrast, scene memory was worse in the mixed-category condition than the single-category condition, $F(1, 19) = 4.33, p < .05$, $\eta^2_p = .19$. This effect also did not interact with encoding duration, $F < 1$.

**Discussion**

Experiment 2 made the scene stimuli more homogenous, ensuring that cortical resources were more limited when encoding four scenes than when encoding two scenes and two faces. Nonetheless we continued to find a mixed-category advantage for face memory but not for scene memory. In fact, scene memory was slightly impaired when scenes were encoded with faces rather than with other scenes, and this deficit reached statistical significance in Experiment 2.

One factor that can produce an asymmetry between faces and scenes is if participants gave greater attentional priority to faces than to scenes when these were mixed together. Consistent with this possibility, Experiment 2 revealed a mixed-category tradeoff between faces and scenes. Unequal priority may originate from an attentional preference for animate stimuli (New, Cosmides, & Tooby, 2007), such as faces (Lavie, Ro, & Russell, 2003). One way in which the preference may have being implemented in Experiments 1 and 2 was by devoting more of the encoding time to faces than to scenes. Participants could use this strategy because stimuli were presented simultaneously, giving them the flexibility to allocate more processing time to faces. In Experiment 3 we aimed to reduce the uneven allocation of encoding time to faces and scenes.

**Experiment 3**

To ensure that participants could not spend disproportionately longer time to encode faces than scenes, in Experiment 3 we displayed the four encoding items sequentially at a pace of 200ms/item. Because each item was presented for a fixed amount of time, participants could not devote more encoding time to faces than to scenes. In addition, we used fast presentation pace to maximize perceptual load (Lavie, Hirst, de Fockert, & Viding, 2004), reducing the likelihood that participants would have excess resources to spare. Using a fast presentation pace also put participants in a reactive mode of encoding. Although 200ms is adequate for people to classify whether an image is a face or a scene (Grill-Spector & Kanwisher, 2005), it leaves little time to voluntarily shift attention to faces (Müller & Rabbitt, 1989). If the lack of a mixed-category advantage for scenes was driven by prioritization of faces, then by limiting uneven allocation of processing time, Experiment 3 should produce a mixed-category advantage for both faces and scenes. On the other hand, if the mixed-category advantage reflected primarily within-category interference for faces, then equating processing time may not alter the asymmetry between faces and scenes.

A second difference between Experiments 3 and the first two experiments was the eccentricity of the stimuli. Instead of presenting stimuli in the four peripheral locations, all stimuli were presented centrally. Two reasons led us to use central presentation in Experiment 3. First, central presentation was necessary to bring performance above chance. Perhaps owing to apparent motion, when the stimuli were presented sequentially in the peripheral locations, memory accuracy was near chance in our preliminary data (mean 57.5%, $N=18$). Presenting stimuli centrally raised the accuracy level.
Second, face and scene processing areas in the ventral stream differ not only in their category selectivity, but also in their central-periphery organization. Whereas face processing overlaps with the processing of centrally presented stimuli, scene processing overlaps with the processing of peripherally presented stimuli (Levy, Hasson, Avidan, Hendler, & Malach, 2001). In Experiments 1 and 2 we presented all stimuli peripherally, which may have exacerbated suppressive interactions among faces that are naturally processed centrally. By presenting all stimuli centrally, Experiment 3 minimized crowding and other suppressive interactions specific to peripherally presented stimuli.

**Method**

**Participants.** Thirty participants (24 females and 6 males with a mean age of 21.2 years) completed Experiment 3.

**Design and procedure.** Faces were randomly chosen from the entire set of 40 faces. The first 20 participants were tested using the same stimuli as those of Experiment 1. That is, scenes were drawn randomly from the entire set of 40, without restrictions to specific sub-categories. To ensure that the results generalized, the last 10 participants were tested using the same stimuli as those of Experiment 2. That is, scenes on a given display always belonged to the same sub-category.

Each participant completed 384 trials. There were equal numbers of “all faces,” “all scenes,” “mixed images with faces tested,” and “mixed images with scenes tested.” On each trial four images were presented at the center of the display at a pace of 200ms/item. The total encoding duration was 800ms, making this experiment comparable to the limited-encoding condition of Experiments 1 and 2. After a blank interval of 1 second, a test image was presented for participants to make a “same/different” judgment. On “same” trials the test image was the same as one of the four encoding images. On “different” trials it differed from all four images.

We probed each encoding serial position equally often, such that on “same” trials the test probe could match the stimulus at each temporal position 25% of the time. On “different” trials the test probe was a different image relative to all encoding stimuli (e.g., a different face relative to the two faces encoded in the mixed-category condition). On these trials it was unclear which temporal position the test probe should be assigned to in the data analysis. Owing to this ambiguity, in most of our analyses we averaged data across all temporal positions.

**Results**

![Sequential presentation](image)

*Figure 6. Results from Experiment 3. Left: Data from participants tested with faces and random scenes. Right: Data from participants tested with faces and sub-category scenes. Error bars show ±1 S.E. of the mean across participants.*

Some participants were tested with scenes drawn randomly (Figure 6, left), others with scenes drawn from a sub-category (Figure 6, right). To examine whether results differed for these two groups of participants, we entered scene materials (random or sub-category) as a between-subject factor in the analysis. None of the factors interacted with scene materials, all ps > .20. Memory was better for the mixed-category stimuli than single-category stimuli, $F(1, 28) = 35.69, p < .001, \eta_p^2 = .56.$
Memory was better for faces than for scenes, $F(1, 28) = 6.91, p < .02, \eta^2_p = .20$. Replicating Experiments 1 and 2, the mixed-category advantage was significantly greater for faces than for scenes, yielding a significant interaction between stimulus materials and category mixing, $F(1, 28) = 12.33, p < .002, \eta^2_p = .31$. Planned contrasts showed that face memory was significantly better in the mixed-category than single-category condition, $t(29) = 6.76, p < .001$. This was not the case for scenes, $t(29) = 1.74, p > .09$.

Although sequential presentation reduced the possibility that participants might allocate greater encoding time to faces, it introduced potential interference among images presented at different serial positions. For example, attending to the first image may exhaust resources for subsequent images, a type of interference akin to the attentional blink (Raymond, Shapiro, & Arnell, 1992). Because faces are less vulnerable to the attentional blink than are other stimuli (Awh et al., 2004), they may have suffered less interference. However, three reasons make us think that sequential interference could not explain the asymmetry in the mixed-category advantage for faces and scenes. First, a reduced susceptibility to the attentional blink for faces should have led to an overall boost in memory for faces in both the single-category and mixed-category conditions. In our data, this was only the case in the mixed-category condition. Second, in our design all four images must be encoded to memory, a presentation condition reminiscent of the “whole report” procedure used in previous attentional blink studies (Nieuwenstein & Potter, 2006). The whole-report procedure is known to produce reduced or no attentional blink.

Finally, if sequential interference contributed to the asymmetry in the mixed-category advantage, then the asymmetry should be most pronounced in temporal positions vulnerable to interference; that is, in the intermediate positions which may be subjected to the attentional blink). This, however, was not the case. Figure 7 plots face and scene memory across the four temporal positions. We performed an ANOVA on stimulus materials, category mixture, and temporal position. The main effect of temporal position was significant, $F(3, 87) = 61.91, \eta^2_p = .68$, revealing the characteristic recency effect in visual working memory (Kumar & Jiang, 2005). However, temporal position did not interact with stimulus materials, $F(3, 87) = 1.93, p > .10$, nor did it interact with category mixing, $F < 1$. The three-way interaction between temporal position, stimulus materials, and category mixing was also not significant, $F(3, 87) = 1.57, p > .20$. These results suggested that sequential interference was unlikely a source for the asymmetry in the mixed-category effect for faces and scenes.

![Figure 7](image)

Figure 7. Results from Experiment 3, separated for images presented at different encoding positions. Error bars show ±1 S.E. of the mean across all 30 participants.

**Cross experiment comparisons**

How did presentation mode influence the results? Here we compared results from Experiment 3 with those of the first two experiments using data from the 800ms duration condition. We entered presentation mode (sequential or simultaneous) as a between-subject factor, and stimulus materials (faces or scenes) and category (single or mixed) as within-subject factors in an ANOVA. This analysis did not reveal significant interaction effects of presentation mode with any other factors, all $Fs < 2.98, ps > .09$. 

1. Visual working memory
2. Cross experiment comparisons
3. Figure 7: Results from Experiment 3, separated for images presented at different encoding positions. Error bars show ±1 S.E. of the mean across all 30 participants.
Combined across all three experiments (using the 800ms encoding condition), we found that face memory was 7.1% better in the mixed-category than single-category conditions, $t(69) = 7.36, p < .001$, effect size Cohen’s $d = 1.77$. Scene memory was 0.3% worse in the mixed-category than single-category conditions, a difference that was not significant, $t(69) = -0.33, p > .50$. If we simply averaged data across faces and scenes, we had a significant mixed-category advantage of about 3.5%, $F(1, 69) = 21.02, p < .001$, $\eta^2 = .19$, an effect comparable in magnitude to that reported by Cohen et al. (2014). The significant interaction with stimulus materials, however, showed that the mixed-category advantage was mainly driven by faces but not by scenes, $F(1, 69) = 42.77, p < .001$, $\eta^2 = .38$.

**Discussion**

Experiment 3 presented faces and scenes sequentially, equating encoding time for faces and scenes. In addition, stimuli were presented centrally, a condition more similar to how participants naturally view faces. The rapid presentation pace also made it difficult for participants to voluntarily attend more to faces than to scenes. Nonetheless we still observed a strong mixed-category advantage for faces but not for scenes. Scene memory was no longer worse when scenes were mixed with faces compared with the all-scenes condition, but unlike face memory, scene memory did not exhibit a significant mixed-category advantage. This finding strengthened the idea that the mixed-category advantage was likely attributable to unique interactions within a category, such as face-specific interference among multiple faces. Although scenes, especially those from the same sub-category (e.g., highway scenes), competed strongly for visual working memory, such competition likely rests on generic factors (e.g., constraints on the number and precision of items that could be remembered), rather than suppressive interactions among multiple scenes.

**Experiment 4**

Findings from the first three experiments raised the question of whether categories other than faces may also be susceptible to category-specific resource limitations. Previous research suggests that faces may be a special category of stimuli, processed by a dedicated, domain-specific recognition mechanism (Kanwisher & Yovel, 2006). All faces share the same basic features and first-order configuration, which may lead to unique interactions among an ensemble of faces, such as the automatic extraction of an “average” face. Mixing faces with other stimuli such as scenes reduces the number of faces on the display and attenuates face-specific interference. On this account, the mixed-category advantage may be a unique, face-specific effect. Alternatively, within-category interference may be a general phenomenon applicable to stimuli distinguished from one another by a spatial rearrangement of the same basic features. Accurate memory for these stimuli requires participants to remember not just the features, but also how the features are spatially arranged to form an exemplar. One category of stimuli that shares basic features is human bodies. The 40 body images used by Cohen et al. (2014), for instance, all depicted the same individual making various sports-related poses. The parts themselves do not distinguish one body image from the other; rather, the global configuration formed by individual parts is a necessary component of memory. It is possible that memory for bodies would also be vulnerable to within-category interference. If this is the case, then the mixed-category advantage should generalize beyond faces.

We have so far tested only faces and scenes. In contrast, Cohen et al. (2014) additionally tested objects and bodies. Their data report did not separate different categories (e.g., they only displayed the average memory for objects and scenes). However, if we assume that faces always show a mixed-category advantage and scenes never do regardless of what other categories they are paired with, Cohen et al.’s data appear to support a mixed-category advantage for faces and bodies, but not for scenes or objects. Experiment 4 directly tested this possibility. We examined face memory, along with memory for two additional categories: bodies and objects. All stimulus combinations were tested in a within-subject design (as opposed to a between-subject design used in Cohen et al., 2014), allowing us to better control for variability across participants.
Method

Participants. Twenty participants (14 females and 6 males with a mean age of 19.4 years) completed Experiment 4.

Stimuli. We used three types of stimuli from Cohen et al. (2014)’s stimulus set: 40 faces, 40 objects, and 40 body images. Faces were the same as those used in Experiments 1-3. Objects were photographs of common objects, such as a bagel, a clock, a button, and a ball. All objects were approximately round in shape. Body images were photographs of an individual in various sports poses (e.g., hitting a golf ball, making a kongfu kick); the head of the individual was not visible. We did not test scenes in Experiment 4 because its inclusion would have made the experiment longer than the maximum IRB approved time of 1 hour.

Procedure and design. The procedure was similar to the limited-encoding condition of Experiment 1. On each trial, participants viewed 4 images presented simultaneously, one in each visual quadrant (Figure 1), for 800ms. After a blank delay of 1,000ms they were presented with the test display. Three of the stimuli on the test display were the same as the stimuli in the corresponding locations on the encoding display; the fourth, enclosed by a red frame, was either the same as the corresponding encoding stimulus (“same” trials, 50%) or a different exemplar (“different” trials, 50%). Participants reported whether the framed image was the same or different from the item in that location before.

The experiment contained 360 trials, divided randomly and evenly into 9 types of trials. Three types of trials were single-category trials: (1) four faces, (2) four body images, or (3) four objects. The other six types of trials were mixed-category trials involving two faces and two objects, two faces and two body images, or two body images and two objects. For each mixture such as faces and objects, the test probe could be either category (e.g., a face or an object) with equal probability. Thus, when face memory was tested, the face may have been encoded with other faces (single-category), with bodies (mixed-category) or with objects (mixed-category). Similarly, when body memory was tested, the body image may have been encoded with other body images (single-category), with faces (mixed-category) or with objects (mixed-category); and so on for object memory, for a total of nine conditions. All types of trials were presented in a randomly mixed order.

Results

Figure 8 displays memory accuracy for the three types of stimuli (faces, bodies, objects), as a function of the category composition on the encoding display.

![Figure 8. Results from Experiment 4. In each panel the left most bar corresponds to the single-category condition. The middle bar is the first type of mixed-category, which included a mixture with faces or bodies. The right bar is the second type of mixed-category, which included a mixture with objects. Error bars show ±1 S.E. of the mean across participants.](image)

We performed a repeated-measures ANOVA using stimulus materials (faces, bodies, or objects) and category mixing (single-category, two types of mixed-categories) as factors. The first type of mixed-category display always included faces or bodies, and the second type of mixed-category display always included objects. Overall accuracy was comparable among the different stimulus...
materials, producing no main effects of stimuli, $F < 1$. The mixed-category conditions showed higher accuracy than the single-category condition, yielding a significant main effect of category mixing, $F(2, 38) = 16.03$, $p < .001$, $\eta^2_p = .46$. Trend analysis showed a significant linear trend in category mixing: accuracy was higher when the target category was encoded with objects than when it was encoded with faces or bodies. The interaction between stimulus materials and category mixing was not significant, $F(4, 76) = 1.79$, $p > .10$. However, the linear trend in the interaction term was highly significant, $F(1, 19) = 8.42$, $p < .009$, $\eta^2_p = .31$. Whereas memory for faces and bodies both exhibited the mixed-category advantage, especially when each was encoded with objects, memory for objects was not greater in the mixed-category condition than the single-category condition. This pattern was supported by follow-up contrasts. The main effect of category mixing was significant for faces, $F(2, 38) = 15.36$, $p < .001$, $\eta^2_p = .45$, and for bodies, $F(2, 38) = 6.58$, $p < .004$, $\eta^2_p = .26$. A significant linear trend of category mixture was found for both types of stimuli: the mixed-category advantage was stronger when faces were mixed with objects than when faces were mixed with bodies, though both were statistically significant ($p < .05$ when comparing all-faces with face+body trials; $p < .001$ when comparing all faces with face+object trials). Similarly, the mixed-category advantage was stronger when bodies were mixed with objects ($p < .002$) than when bodies were mixed with faces ($p > .10$).

Object memory, in contrast, was unaffected by category mixing, $F < 1$.

A further follow-up experiment (N=20) reproduced the finding that memory for bodies showed a mixed-category advantage. This experiment used scenes, objects, and bodies as stimulus materials, essentially replacing the face stimuli used in Experiment 4 with scenes. Scenes on a given display always belonged to the same sub-category. This experiment showed that memory for bodies was better when the body images were encoded with scenes (83%) or with objects (83%), relative to the single-category condition (76%; $p < .001$). In contrast, memory for scenes was comparable whether scenes were encoded with other scenes (71%) or with objects (71%) or bodies (70%; $F < 1$). Similarly, memory for objects was comparable regardless of whether objects were encoded with other objects (83%) or with scenes (85%) or bodies (81%; $p > .10$).

**Discussion**

Experiment 4 generalized the mixed-category advantage in visual working memory from faces to images of human bodies. This finding rules out one account of the mixed-category advantage: that it might have been driven by face-specific computations. This experiment also showed that scenes were not the only category that was insensitive to category mixing. Object memory also failed to improve when objects were encoded with other categories (including faces, scenes, and bodies) relative to the single-category condition. As discussed in the General Discussion section, these data constrain theoretical interpretations of the mixed-category effect.

**General Discussion**

Although working memory is a core component of intelligence and other central processes (Miyake et al., 2000; Shipstead et al., 2015), sensory and perceptual regions of the brain also contribute to working memory (Eimer, 2015; Harrison & Tong, 2009; Ranganath et al., 2004). Evidence for the involvement of category-specific cortical resources comes from the mixed-category advantage: working memory for a mixed display of faces and scenes is superior to that for just faces or scenes (Cohen et al., 2014). The present study extended conditions under which a mixed-category advantage was observed. Our findings also raised challenges to previous interpretations of this effect.

Replicating Cohen et al. (2014) we found that mixing faces and scenes in encoding conveyed an overall advantage to working memory performance. In addition, we generalized this finding to conditions in which participants had adequate time to encode all stimuli to memory. Removing encoding limitation as a possible account establishes the mixed-category advantage as a characteristic of visual working memory.

Although we observed an overall mixed-category advantage, one novel finding of the present study is that the effect was not equally strong for different categories of stimuli. In the first three
experiments, we found a clear advantage to face memory when faces were encoded with scenes rather than with other faces. The mixed-category advantage is therefore a robust phenomenon in face memory. However, we found no advantage to scene memory when scenes were encoded with faces rather than with other scenes. This finding held regardless of presentation mode (simultaneous or sequential), presentation duration (limited or self-paced), or stimulus eccentricity (peripheral or central). Given the reduced overlap in cortical processing for a mixed display of scenes and faces compared to a display of all scenes, the scene data posed difficulties for the cortical resource theory. A modified version of the theory, in which the mixed-category advantage correlates negatively with the availability of cortical resources for each category, also fails to account for the entire dataset. This is because restricting scene stimuli to a specific sub-category should have restricted cortical resource for scenes. Yet replacing scenes with faces did not produce a mixed-category advantage for scenes.

The scene data cannot be dismissed as a null result. Combined across all participants, the observed statistical power for detecting the mixed-category advantage was 1 for faces ($\eta_p^2$ was 0.44). Our sample size should have been adequate for detecting a comparable effect for scenes. In addition, in Experiment 2 in which sub-category scenes were used, mixing scenes with faces did not just yield a null result – it significantly impaired scene memory. Finally, the asymmetry between scenes and faces was supported by statistically significant interactions between stimulus materials (faces or scenes) and category mixing (single or mixed). The empirical data leave few doubts that the mixed-category advantage was asymmetric.

The final experiment tested two additional categories, bodies and objects, and showed that the mixed-category advantage was not specific to faces. These data leave open the source of asymmetry between faces and bodies on one hand, and scenes and objects on the other. Faces and bodies differ from objects and scenes in at least two ways. First, the face and body stimuli used in our study required people to remember not just individual parts, but also their spatial configuration. This type of configural memory may be highly limited, resulting in large interference among multiple stimuli from the same category. The interference may arise from face averaging or mis-combination of parts. Replacing two of the stimuli with another category is likely to reduce the within-category interference. In contrast, scenes and objects used here often have unique features that distinguish one exemplar from another. Because configural memory is needed for both faces and bodies, but not for objects, this account can also explain why the mixed-category advantage for faces and bodies was weaker when mixed with each other, than when mixed with objects.

Alternatively, it could be argued that the mixed-category advantage may be stronger for animate stimuli than inanimate stimuli because faces and bodies draw a disproportionately larger share of generalized central resources than do inanimate stimuli (New et al., 2007). Such an uneven allocation of generalized central resources would result in stronger encoding of animate compared to inanimate categories, thus yielding a mixed-category advantage for animate stimuli, but no advantage or a mixed-category deficit for inanimate stimuli. Experiment 3 attempted to control for uneven allocation of resources through the sequential presentation of items, where each item was presented for the same amount of time, equating encoding duration. While not a remedy for all forms of unequal attention allocation, the fact that the results were virtually identical to experiments using simultaneous presentation argues against resource allocation as a full account of our results.

Asymmetric interference in working memory has also been observed previously for visual and verbal materials (see Morey, Morey, van der Reijden, & Holweg, 2013 for a review). For example, Morey et al. (2013) presented participants with visual arrays of colors and spoken digits and probed memory for either the visual or verbal materials. Compared with conditions in which either visual or verbal materials were omitted (single-task), verbal memory was relatively intact when a concurrent visual memory load was imposed, but visual memory was impaired when a concurrent verbal memory load was imposed. In our view, the visual/verbal asymmetry is likely due to the presence of a phonological loop (Baddeley, 2012), which buffers verbal memory in the presence of visual interference. In contrast, visual materials do not have access to a comparably shielded buffer owing to the strong reliance on central resources (Makovski et al., 2006; Vogel, McCollough, & Machizawa,
2005). It is therefore unlikely that the same mechanisms underlie the visual/verbal asymmetry and the face/scene asymmetry. Nonetheless, both patterns of data suggest that a consideration of just central, domain-general resources is inadequate in accounting for how multiple categories of stimuli are retained in working memory.

Our data have implications for how domain-specific cortical resources constrain visual working memory. Until now the cortical resource theory has emphasized primarily the degree of cortical overlap in the ventral processing streams between categories. When the overlap is low, as it is for face processing and scene processing, competition for domain-specific resources is lessened because fewer stimuli in each category compete for these resources, yielding a mixed-category advantage. This theory, however, does not fully explain why an increase in cortical resource has a negligible effect on scene (or object) memory. Neither can the asymmetry between faces and scenes be explained by a difference in memory capacity for faces and scenes. Though people can remember a large number of natural scenes in long-term memory (Konkle et al., 2010), working memory capacity for scenes within a sub-category is limited. In fact, our pilot data (Figure 4) showed that it was more difficult to remember several scenes from the same sub-category than to remember several faces.

To account for the data presented here, it may be necessary to consider three sources of constraints to visual working memory: 1) allocation of domain-general central resources; 2) domain-specific resources; and 3) within category interaction among stimuli. First, domain-general resources include the focus of attention (Cowan, 2001; or “central executive”, Baddeley, 2012), and/or a generic limit in the number of objects that one can remember (Zhang & Luck, 2008). The allocation of general resources typically depends on behavioral goals: categories of stimuli that are more relevant to the current task will likely receive greater resource allocation. Second, category-specific cortical resources are also involved in representing stimuli from specific categories. Memory performance suffers either because memory load exceeds the generic memory limits (e.g., there are too many objects to remember), or because of category-specific interference from stimuli in the same category. Third, within-category interference may be most salient when the different exemplars share the same features or parts. Such similarity may yield automatic averaging of the different exemplars, and it also places strong demand on configural memory. Working memory is in part domain specific because it is limited by within-category interactions. However, not all categories exhibit strong within-category interactions. The similarity of features and holistic processing of faces permits an automatic extraction of an average face, at the cost of representing individual faces (de Fockert & Wolfenstein, 2009). This within-category interference is reduced when faces are encoded with scenes rather than with other faces. The release from within-category interference contributes to the mixed-category advantage. Because both faces and bodies exert demands on configural memory, mixing these two categories yielded less advantage relative to when they were mixed with objects or with scenes.

Finally, we note that within-category interactions for faces are not always suppressive. Memory for highly similar faces, such as face morphs generated from the same identity, is more precise than memory for different faces (e.g., face morphs generated from different identities; Jiang, Lee, Asaad, & Remington, 2015). This finding is the opposite of what the cortical resource theory would predict: substantial overlap in cortical resources for highly similar faces enhances, rather than interferes, with visual working memory (see also Lin & Luck, 2009, for color memory, and Sims, Jacobs, & Knill, 2012, for orientation memory). This is another case in which the nature of within-category interaction, rather than the size of the resource pool, determines whether similarity helps or hinders memory (Sims et al., 2012).

Conclusion

Our study both extended the cortical resource theory and provided important constraints. We showed that the mixed-category advantage was not limited to brief presentation conditions, suggesting that it was a characteristic of working memory rather than just perception. We also showed that the mixed-category advantage was highly robust for faces and bodies. However, scene memory was not enhanced when scenes were encoded with faces rather than with other scenes. This
was observed not only for scenes drawn from multiple categories, but also for scenes from a specific sub-category. Presenting the stimuli sequentially rather than simultaneously did not significantly alter the results. The lack of a mixed-category advantage extended to objects. We propose that an increase in the total pool of cortical resources can facilitate working memory, but only for categories that exhibit strong within-category interference. Such interference may come from automatic averaging of multiple exemplars or the difficulty in remembering spatial configurations of features and parts. Future studies that test memory for a broad range of categories are needed to delineate the nature of within-category interactions in visual working memory.

Acknowledgements

We thank Sha Li and Simon Ozbek for help with data collection. Correspondence should be directed to Yuhong Jiang, 75 East River Road, S251 Elliott Hall, Department of Psychology, University of Minnesota, Minneapolis, MN 55455. Email: jiang166@umn.edu.

References


Visual working memory


