**The Asymmetric Mixed-Category Advantage in Visual Working Memory:
An Attentional, Not Perceptual (Face-Specific) Account**

Journal of Experimental Psychology: Human Perception and Performance (*in press*).

Rotem Avital-Cohen & Nurit Gronau

Department of Psychology & Cognitive Science Studies, The Open University of
Israel, Raanana, Israel 43107

Word Count: 11,840

Authors' Note

Some of the experiments appearing in the manuscript were presented at the Vision Sciences Society meeting (2019) and the annual meeting of the Psychonomics Society (2019). We thank Anna Izoutcheev, Sharon Yakim and Moshe Sheinfeld for data collection. We thank Tal Makovski for his helpful thoughts and advice. This research was supported by the Israel Science Foundation (grant No. 1622/15) to Nurit Gronau.

Correspondence concerning this article should be addressed to:

Nurit Gronau

Department of Psychology & Cognitive Science Studies, The Open University of Israel,
1 University Rd., Raanana, 4353701, PO box 808, Israel.

e-mail: nuritgro@openu.ac.il

Abstract

The mixed-category advantage in visual working memory refers to improved memory for an image in a display containing two different categories relative to a display containing only one category (Cohen et al., 2014). Jiang, Remington, et al. (2016) found that this advantage characterizes mainly faces and suggested that face-only displays suffer from enhanced interference due to the unique configural nature of faces. Faces, however, possess social and emotional significance that may bias attention toward them in mixed-category displays at the expense of their counterpart category. Consequently, the counterpart category may suffer from little/no advantage, or even an inversed effect. Using a change-detection task, we showed that a category that demonstrated a mixed-category disadvantage when paired with faces, demonstrated a mixed-category advantage when paired with other non-facial categories. Furthermore, manipulating the likelihood of testing a specific category (i.e., changing its task-relevance) in mixed-category trials, altered its advantaged/disadvantaged status, suggesting that the effect may be mediated by attention. Finally, to control for perceptual exposure factors, a sequential presentation experimental version was conducted. Whereas faces showed a typical mixed-category advantage, this pattern was again modulated (yielding an advantage for a non-facial category) when inserting a task-relevance manipulation. Taken together, our findings support a central resource allocation account, according to which the asymmetric mixed-category effect likely stems from an attentional bias to one of the two categories. This attentional bias is not necessarily spatial in its nature, and it presumably affects processing stages subsequent to the initial perceptual encoding phase in working memory.

Keywords: visual working memory, attentional bias, mixed-category effect, face perception.

Public Significance Statement

The present study investigated working memory (WM) for real-world images. People tend to remember better images taken from different categories than from the same category, however, this pattern characterizes only certain types of stimuli (e.g., faces). Here we show that the beneficial memory for faces is not a category-specific effect (e.g., due to their holistic perception). Rather, it is an outcome of an attentional bias to faces, on the expense of memory for other categories. When measuring WM capacity, then, one needs to take into account potential biases, that may occur at the encoding or the retention phase of WM.

Visual working memory (VWM), as several other cognitive functions, is a limited-capacity system. Whether its capacity is limited by the number of the to-be-retained items (e.g., Awh et al., 2007; Cowan, 2001, 2010; Luck & Vogel, 1997; Vogel et al., 2001), their complexity (as determined by the amount and type of features within each item, e.g., Alvarez & Cavanagh, 2004) or both (Fougnie et al., 2010; Olson & Jiang, 2002) is yet controversial. One of the most common tools used to study and measure VWM is the *change-detection* task (Luck & Vogel, 1997; Phillips, 1974; Rouder et al., 2011; for a review see Rensink, 2002), in which participants judge whether a stimulus in the test display – either a simple feature or a more complex image – is the same or different than a stimulus that was just seen within the encoding display. Often, the encoding display contains several to-be-remembered items, or a cluttered scene, such that items perceived and maintained in working memory compete for processing resources.

One of the factors affecting competition, aside from stimulus set size (Luck & Vogel, 1997; Pashler, 1988; Vogel et al., 2001; Wilken & Ma, 2004), is the degree of similarity among items. The role of visual similarity in perceptual and attentional processes has long been documented, as for instance, when impaired visual search performance is observed if attention is attracted to irrelevant distractors that possess features similar to that of the target (Duncan & Humphreys, 1989; Jacobs, 1986; see also Roper et al., 2013). Target-distractor similarity, in this case, directly influences the efficiency of target selection within the visual search task. Competition among stimuli due to their perceived visual (or at times conceptual) similarity, has additionally been found to affect long-term memory processes (Anderson et al., 1994; Baddeley, 1966; Konkle et al., 2010) and working memory processes (Cohen et al., 2014; Mueller et al., 2003; Viswanathan et al., 2010; Wong et al., 2008). In accordance with the visual search findings, results typically reveal that the greater the similarity among items, the greater the impairment in memory performance due to enhanced item-to-item competition.

Note, however, that there are also some opposing results that suggest that under certain circumstances visual similarity may in fact facilitate memory performance (Jiang, Lee, et al., 2016; Lin & Luck, 2009; Shapiro & Miller, 2011).

A recent account for the role of visual and/or conceptual similarity among real world images in visual working memory was suggested by Cohen et al. (2014), who demonstrated a ‘mixed-category advantage’ with the change-detection task. The mixed-category advantage refers to improved memory for an image in a display that includes items taken from several categories relative to a display composed of only single-category items (see, e.g., Wong et al., 2008). Thus, for instance, memory for a specific face image is improved when it appears among a face and two other scenes, than among three other faces. Using behavioral as well as fMRI methods, Cohen et al. suggested that visual working memory is constrained by cortical representational resources. That is, stimuli drawn from the same category are represented by similar neural populations within the visual cortex and therefore strongly compete with each other; whereas stimuli drawn from different categories are represented by different neural populations, hence they only weakly compete with each other. Moreover, the magnitude of the mixed-category benefit, according to Cohen et al., is predicted by the amount of separation between the neural response patterns within occipitotemporal cortex. Thus, presenting faces and scenes (characterized by largely distinct neural activation patterns) yields a large mixed-category advantage, while the presentation of objects and scenes (characterized by large neural activation overlap) yields little or no mixed-category benefit. Importantly, in their study, Cohen et al. collapsed across the categories presented when computing the mixed-category advantage effect. Namely, mixed-category trials in which change detection was performed on one category or another were averaged and compared to single-category trials from both types of stimuli, again averaged across categories.

In a follow-up study carried out by Jiang, Remington, et al. (2016), however, a very different account was suggested for the mixed-category effect. In contrast to Cohen et al.'s study, Jiang, Remington, and colleagues (2016) examined the effects of mixed- vs. single-category trials within each type of category separately, enabling a more refined analysis of the mixed-category benefit. Using behavioral measures, they revealed an interesting asymmetric pattern of results: while a mixed category advantage was found for faces and bodies (when paired with other categories), no such advantage was observed for scenes and objects. Thus, for example, faces showed a mixed-category advantage when paired with scenes, but scenes failed to show an advantage when paired with the very same faces, under the very same presentation conditions. Based on this pattern of results Jiang et al. suggested that the mixed-category advantage cannot be explained by some generalized mechanism of competition for cortical resources; rather, the authors suggested a category-specific interference account that characterizes only a *subset* of category types. Specifically, according to the authors, faces and bodies are subjected to strong within-category interference due to the unique configural or holistic nature of their processing. Due to this strong interference, or within-category competition, such categories benefit from mixed- relative to single-category displays. Other categories, however, whose visual attributes do not elicit such strong within-category interference do not exhibit this benefit (e.g., scenes, objects). Thus, while some categories are characterized by a strong inter-category competition, others are not (or are less so), forming an inherent asymmetry in the mixed-category effect pattern. Note that the assumption of the holistic- and/or configural-related attributes of faces playing a central role in the mixed-category advantage may be supported by the fact that such an effect was shown to be eliminated when using inverted faces (Wong et al., 2008).

Yet faces differ from other types of categories not only in their structural or perceptual properties, but also in their social, emotional and biological significance. Due to their

importance in everyday life, faces (and perhaps bodies) tend to capture and/or engage attention (e.g., Crouzet et al., 2010; Langton et al., 2008; Ro et al., 2007; Ro et al., 2001; Sato & Kawahara, 2015; Theeuwes & Van der Stigchel, 2006) ,potentially depriving other categories present in the display from processing resources required for their encoding and maintenance in VWM. Consequently, any mixed-category benefit within these latter categories (which are paired with the faces) may be dramatically reduced or cancelled out. Put differently, the asymmetric mixed-category benefit may be independent of the visual attributes that characterize a particular set of categories. Rather, it may result from an unbalanced distribution of processing resources during memory encoding and/or retention (i.e., the *central resource allocation* account, see Jiang, Remington, et al., 2016¹). In such case, the asymmetry in memory performance observed among different categories may be the outcome of domain-general capacity limits in VWM (e.g., Cowan, 2001, 2010; Morey & Bieler, 2013; Rouder et al., 2008), rather than domain-specific (category-selective) visual properties.

The present study, therefore, aimed to thoroughly investigate the central capacity limitation account of the asymmetric mixed-category effect. To this end, three lines of studies were designed. The first (Experiments 1-2) replicated Jiang, Remington, et al.'s finding of an asymmetric mixed-category effect when pairing faces with various types of scene categories. In order to improve within-category similarity factors among categories, we used a more controlled set of stimuli than the one used in previous studies. The second line of research (Experiment 3) tested the *specificity* of this asymmetric effect by using *new* pairings of categories that were allegedly characterized by no, or little, inter-category competition (according to Jiang, Remington, et al., 2016). Hence, these categories were not expected to

¹ Jiang, Remington, and colleagues (2016) tried to control for some attentional factors that may affect stimulus perception, by presenting stimuli serially rather than simultaneously. We will refer to the serial presentation manipulation and its potential implications for an central resource allocation account in Experiments 4 and 5, and in the General Discussion.

exhibit a mixed-category advantage since they were not characterized by a holistic, or a configural nature. Once demonstrating that the asymmetric effect is generalized to non-facial categories, a third line of research (experiments 4-5) manipulated *attentional distribution* across pair categories, in an attempt to overcome a potentially inherent bias of attention to one of the two categories, at the expense of the other. That is, we modified the relative attentional distribution across the two categories, while examining these attentional effects on the asymmetric mixed-category effect.

In a series of change-detection experiments, participants made same/different judgments on one of four encoded images that belonged to either one or two categories. Both raw accuracy rates and sensitivity (d') measures were computed in order to assess working memory performance. As the pattern of results was very similar in the two measures, and since the d' measure better controls for guesses and response-biases, only sensitivity results are reported here for all experiments.

Experiment 1: Pairing Faces and Scenes

The first experiment aimed to replicate Jiang, Remington, et al.'s (2016) Experiment 2, which was a follow-up on Cohen et al.'s (2014) original study. In this study, images of faces and scenes were presented in same- or mixed-category trials, and the mixed category advantage was observed only among the facial category. Note that whereas in Cohen et al. scene images within each trial were randomly selected from a wide range of basic-level categories (such as cities, oceans, highways etc.), in the Jiang et al. study scene stimuli within each trial were narrowed to include only images taken from the same basic-level category (e.g., four city images in the single-category condition; two ocean images accompanied by two faces in the mixed-category condition). This was done in order to reduce the within-category variability of the scenes (at the trial level), thus equating their inter-category interference level to that of the faces. While Jiang et al.'s experiment contained two encoding duration conditions (800 ms

and a self-paced duration, manipulated in separated blocks), in the present experiment only a 800 ms encoding duration was utilized.

Method

Participants. Twenty participants (15 females, ages 20-39) took part in the experiment for course credit or for payment. Sample size was determined based on the number of participants tested in Jiang, Remington, et al.'s (2016) Experiment 2. All had normal or corrected-to-normal vision. The experiment (as well as all of the following experiments) has been approved by the Israel's Open University ethical review board committee. All participants signed a consent form indicating that participation was voluntary and that they could withdraw from the experiment at any time without penalty.

Apparatus. Participants were seated in a dimly lit room, with their heads placed on a chin rest that was used to stabilize their viewing distance at 60 cm from a 23.5-in. LCD monitor (1920 × 1080 px; 120 Hz) on which stimuli were displayed. Stimuli presentation and data collection were controlled using OpenSesame software (Mathôt et al., 2012).

Stimuli. The 40 scenes images were identical to those used by Cohen et al. (2014) and by Jiang, Remington, et al. (2016), containing 10 exemplars from each of four basic-level scene categories: city images, highways, mountains, and oceans. The faces set was taken from the copyright-free FEI Face Database (Thomaz, 2012, see *frontalimages_manuallyaligned_part1* and *part2*). It contained 22 women images, and an additional 18 men facial images, that were chosen to create a stimulus set as similar to the one used in the two previous studies (Cohen et al., 2014; Jiang, Remington, et al., 2016). Facial expressions were similarly matched to those of the Cohen/Jiang et al. studies (with most faces smiling). All faces were cropped to remove hair contour, and as in previous studies, all

images – scenes and faces – were presented in grayscale (see Appendix for image examples of the two categories).

Experimental Design and Procedure. Each trial began with 300 ms fixation display of a $0.55^\circ \times 0.55^\circ$ central red fixation point, followed by an 800 ms encoding display of four images, each subtending $6^\circ \times 6^\circ$. The centers of the four images formed an imaginary $15.4^\circ \times 7.5^\circ$ rectangle, which was centered at fixation. The encoding display was followed by a 1,000 ms blank retention interval, after which the test display appeared. The test display was similar to the original encoding display in containing four images, three of them identical to the encoding images (and presented at the very same locations), and an additional image surrounded by a $8^\circ \times 8^\circ$ red frame (of 0.13° width) that was either identical to the encoding image (50% of trials), or was different from it. In the latter case, a different exemplar from the same basic-level category appeared (with no re-sampling of images within a trial). The location of the test item was determined randomly. Participants were asked to judge whether the test item was the same or different from the item presented at its location at the encoding phase, by pressing the "j" or "k" keys, respectively. The test display was presented until response execution. Each response was followed by an auditory feedback. The inter-trial interval was 1,000 ms. Figure 1 illustrates an example of a trial sequence.

The experiment consisted of 320 trials divided into 10 blocks of 32 trials each, which were preceded by 20 practice trials. Half of the experimental trials contained a single-category display in which all four images belonged to the same basic-level category (e.g., four faces or four mountains), and half were mixed-category display trials in which the display contained two faces and two scenes images taken from the same basic-level category (e.g., city images). The specific stimuli were assigned randomly to the four different locations. Overall, the experiment included four different test-item conditions, resulting from the joint manipulation of the display-type and the categorical factors: 80 trials contained a single-category display of

four faces, 80 trials contained a single-category display of four scenes, 80 trials contained a mixed-category display with a face as the test item, and the remaining 80 trials contained a mixed-category display with a scene as the test item. Within each of the conditions, half of the trials contained a "same" test display and half contained a "different" test display. All trial types were randomly intermixed throughout the experiment. The experiment lasted approximately 40 minutes, including oral instructions, practice and short breaks.

Figure 1. An example of a trial sequence in Experiment 1, containing a mixed-category display for which a “Different” response was required. For the scene stimuli, although there were four different types of scene categories, items within each trial were narrowed to include only images taken from the same basic-level category (e.g., two highways in the present example).

Results and Discussion

Overall accuracy levels in the change-detection task were typical to this type of task (mean = 0.73, SD = 0.08). Figure 2 presents d' means for the four experimental conditions. A Category (faces vs. scenes) $\times$ Display-type (single- vs. mixed-category) repeated measures analysis of variance (ANOVA) performed on the d' means revealed a significant main effect of category, $F(1, 19) = 16.32, p < .001, \eta_p^2 = .46, BF_{10} > 100^2$, stemming from an overall higher sensitivity to faces than to scenes. A significant main effect was additionally obtained with the display-type factor, $F(1, 19) = 4.41, p = .049, \eta_p^2 = .19, BF_{10} = 0.75$, resulting from greater sensitivity in the mixed-category display than in the single-category display (though this result was inconclusive according to Bayesian analysis). Of most interest, however, was the significant interaction between category and display-type, $F(1, 19) = 10.89, p = .004, \eta_p^2 = .36, BF_{10} = 26.82$, which was clarified by further analyses of the simple effects: memory for faces was better in the mixed-category display than in the single-category display, $t(19) = 3.76, p = .001$, Cohen's $d = 0.84^3$, $BF_{10} = 28.47$, while memory for scenes did not significantly differ between the two display types, $t(19) = -1.22, p > .1$, Cohen's $d = 0.27$, $BF_{10} = 0.45$ (inconclusive according to Bayesian norms). The mixed-category advantage was observed, then, only for faces, replicating Jiang, Remington, et al.'s main findings (2016) of an asymmetric mixed-category advantage when using face and scene images. Beyond the

² We conducted a Bayesian analysis (e.g., Dienes, 2011; Jeffreys, 1961) for the Anova, in which the Bayes factor for each of the effects ($BF_{inclusion}$) compares a model that contains the effect with equivalent models stripped of the effect. The prior values for the Bayesian Anova were 0.5 and 1 (for the r scale fixed and random effects, respectively). For the follow-up simple effects of the interaction, the Bayes factor (BF_{10}) represented the likelihood ratio of a mixed category advantage. The prior value for these Bayesian contrasts was 0.707 (i.e., the default value of JASP software). According to widely accepted benchmarks, BF_{10} values of 10 or greater indicate a strong effect, $3 < BF_{10} < 10$ suggests a moderate effect, $0.33 < BF_{10} < 3$ implies an inconclusive result, and $BF_{10} < 0.33$ indicates the absence of an effect (i.e., a clear support for H_0).

³ Cohen's d effect size is defined as the standardized mean difference between the two conditions. According to common rules of thumb, d equaling 0.2, 0.5, and 0.8 stand for small, medium, and large effects, respectively (see Cohen, 1988).

importance of the reliability of these findings, the replicated results imply that the novel set of face images used in the present study behaves similarly to the facial dataset used in previous studies, and the two facial sets are thus highly comparable. The present results, then, corroborate on previous ones in demonstrating a mixed-category effect uniquely characterizing facial stimuli, but not scene stimuli.

Figure 2. Mean d' for each display-type condition, within the different category conditions in Experiment 1. Error bars represent ± 1 *SE* of the means.

Note that in Experiment 1, although the level of homogeneousness among scene images was controlled for at the trial level, the two categories— scenes and faces – were yet unbalanced, as scenes were divided into four highly distinct basic-level categories. That is, while faces were sampled from a set of 40 stimuli, scenes in each trial were sampled from a set of merely 10 stimuli (that formed each of the basic-level scene categories). This raised the probability for the joint appearance of specific image exemplars within the same trial, on the one hand, while increasing the overall distinctiveness of scene types across trials, on the other hand. The use of four narrowed categories, rather than one, within the scene superordinate category may

have therefore eased the process of learning the scene images and familiarizing with them. This familiarization, in turn, may have attenuated stimulus competition within each trial, reducing the likelihood of a mixed-category advantage among the scenes. Experiment 2 was designed to address this issue and control for the unbalanced sampling properties of the two categories.

Experiment 2: Pairing Faces and Highways/Clocks

In order to reduce conceptual and visual differences among the various scene types and render the scene category more comparable to the faces, in Experiment 2 only one basic-level scene category was used. In addition, we tested the mixed-effect advantage by examining an additional basic-level object category paired with faces. Note that Jiang, Remington, et al. (2016) found an asymmetric pattern of mixed-category advantage, resembling the one found with scenes, when pairing objects with faces. The superordinate category of objects in their study, though, included a wide range of basic-level categories, such as clocks, buttons, rings, etc. (sampled randomly within and across trials). To control for within-category homogeneity among the different categories, in Experiment 2A, therefore, 40 *highway* images were chosen to represent scenes, all depicting a highway viewed from the same viewpoint and containing a roughly identical configuration and semantic content. In Experiment 2B, 40 *clocks* were chosen to represent objects, consisting of a fixed global configuration (i.e., round contour) as well as similar first-order spatial relations among the clocks' constituents (e.g., digits, hands)⁴. To the extent that feature-overlap (and/or spatial

⁴ It was previously proposed that the demands on configural memory underlie the increased interference causing the mixed-category advantage effect among faces (but not among other stimulus categories, e.g., Jiang, Remington, et al., 2016). Despite the obvious social and biological differences between faces and clocks, and in spite of the fact that clocks may not be processed holistically or configurally as faces, the two types of categories share some global properties (e.g., round contour, fixed spatial configuration), and thus may behave similarly in the context of the present study. Indeed, due to these shared properties, clocks were used as control for face images in several visual working memory studies (e.g., Cheung & Gauthier, 2010; Curby & Gauthier, 2007; Meinhardt-Injac et al., 2013; Wong et al., 2008).

configuration) may account for image competition, the present experiments should provide improved conditions for testing the mixed-category effect within each of the categories.

Method

Participants. Thirty participants (24 females, ages 20-40) took part in Experiment 2A, and an additional 30 participants (23 females, ages 18-39) took part in Experiment 2B. The participants received course credit or payment, and all had normal or corrected-to-normal vision. Sample size was determined based on a post-hoc power analysis of the mixed-category effect observed with faces in Experiment 1, which equaled 0.95. Since faces have constantly demonstrated a mixed-category advantage (here and in previous studies), and due to the use of additional categories that may not exhibit such a robust effect, we assumed a more conservative effect size than the one found in Experiment 1 (Cohen's $d = 0.84$). An *a-priori* power analysis was conducted with the G*Power tool (Faul et al., 2009), using an effect size of Cohen's $d = 0.70$ and an alpha of 0.5. This analysis yielded a necessity of about 30 participants (given a power of 0.95). From now on, then, 30 participants took part in each of the experiments in this paper.

Apparatus, Stimuli and Experimental Design. Stimulus display and procedure were similar to these of Experiment 1 with the exception that all scene images in Experiment 2A depicted highways, and all object images in Experiment 2B depicted clocks. As mentioned above, the highways of Experiment 2A were pictured from the same viewpoint (i.e., extending from the observer and onwards) and represented a relatively similar overall 'gist'. There were, however, differences among images with respect to road brightness, the exact surrounding scenery (e.g., mountains, fields) or the specific objects embedded within the scene (e.g., cars, bridges, signposts). All highway images were taken from the same database as the original scene images. The clocks in Experiment 2B were all analog clocks, round in shape, containing either Arabic or Roman numerals, and showing the time 10:10. The clocks

differed in their frame, dial, and hands (see Appendix for image examples of the categories used in the two experiments). Clock images were taken from the internet using Google images search and were modified where needed. All other aspects of the two experiments were identical to Experiment 1.

Results and Discussion

The mean accuracy level in the change detection task was 0.68 (SD = 0.10) in Experiment 2A and 0.69 (SD = 0.10) in Experiment 2B. Figure 3 presents d' means for the four experimental conditions in each of the experiments. We first describe the results of Experiment 2A (faces versus highways) and then of Experiment 2B (faces versus clocks). A Category (faces vs. highways) $\times$ Display-type (single- vs. mixed-category) repeated measures analysis of variance (ANOVA) performed on the d' means of Experiment 2A revealed a significant main effect of category, $F(1, 29) = 28.35, p < .001, \eta_p^2 = .49, BF_{10} > 100$, resulting from a higher sensitivity for faces than for highways. There was no main effect of display-type, $F(1, 29) = 2.06, p > .1, \eta_p^2 = .07, BF_{10} = 0.32$. Again, of main interest was the robust interaction between the category and display-type, $F(1, 29) = 26.14, p < .001, \eta_p^2 = .47, BF_{10} > 100$, which was clarified by further analyses of the simple effects: memory for faces was better in the mixed-category display than in the single-category display, $t(29) = 4.56, p < .001$, Cohen's $d = 0.83, BF_{10} > 100$, while there was no such mixed-effect advantage with the highways. In fact, memory for the latter category was significantly worse in the mixed-category display than in the single-category display, $t(29) = -2.68, p = .012$, Cohen's $d = 0.49, BF_{10} = 3.84$.

The repeated measures ANOVA performed on the d' means of Experiment 2B (faces versus clocks) demonstrated once again a significant main effect of category, $F(1, 29) = 132.90, p < .001, \eta_p^2 = .82, BF_{10} > 100$, stemming from better memory for faces than for

clocks. In addition, a main effect of display-type was observed, $F(1, 29) = 11.93, p = .002, \eta_p^2 = .29, BF_{10} = 4.63$, yet it was qualified by a category $\times$ display-type interaction, $F(1, 29) = 20.46, p < .001, \eta_p^2 = .41, BF_{10} > 100$. As in the faces-highways experiment (2A), an increased memory for faces was obtained within the mixed- relative to the single-category display, $t(29) = 4.95, p < .001$, Cohen's $d = 0.90, BF_{10} > 100$. No such mixed-category advantage was observed with the clocks (if anything, a non-significant trend was seen in the opposite direction), $t(29) = -1.93, p = 0.063$, Cohen's $d = 0.35, BF_{10} = 0.99$.

Figure 3. Mean d' for each condition in (a) Experiment 2A and (b) Experiment 2B.

Error bars show ± 1 SE of the mean across participants.

It appears that despite the use of homogeneous sets of scenes and objects that were similar to the set of faces in terms of their overall structure and sampling properties, a mixed-category advantage was again demonstrated only for faces, replicating Jiang, Remington, et al.'s (2016) findings. Although it is hard to quantify the precise degree of homogeneousness among the different sets, these results suggest that within-category similarity (i.e., homogeneousness) cannot, in itself, account for a category's disposition to the mixed-category advantage. Faces, thus, may indeed invoke a unique pattern of memory performance when paired with other non-facial categories.

And yet, in order to reach a firm conclusion with regards to the uniqueness of faces, one should rule out the existence of a mixed-category advantage among *non*-facial categories, when paired with each other⁵. To the extent that the mixed-category advantage is an outcome

⁵ Jiang, Remington, et al. (2016) mentioned a follow-up experiment that included scenes, objects and bodies, in which no mixed-category effects were observed either for scenes or objects when paired with each other (p. 1408). Here we tested these two categories against each other when using narrowed, basic-level categories, rather than wide superordinate categories as the ones used in the Jiang, Remington, et al. and in the Cohen et al. (2014) studies.

of the specific visual properties and/or the unique type of processing that characterizes faces (and perhaps bodies) – no effect should be observed when testing the joint interaction of objects and scenes. If, however, the mixed-category advantage is independent of specific visual attributes and results mainly from an unbalanced distribution of processing resources during memory encoding and/or retention (i.e., the central resource allocation account), an asymmetrical pattern of results similar to the one obtained with the faces may be observed with non-facial categories. Note that some hint for a central resource allocation account comes from these two experiments (2A and 2B), in which the highways and the clocks categories in fact showed a trend for a reversed mixed-category effect. It may be the case that this tendency for a mixed-category *disadvantage* stems from greater attention resources being deployed to the facial images in the mixed-category trials, at the expense of the scenes (see similar findings in Jiang, Remington, et al., 2016, Experiment 2). Although we did not have a clear hypothesis concerning the specific attentional distribution pattern among objects and scenes, Experiment 3 tested whether an asymmetrical mixed-effect may arise when conjoining such categories within the same visual display.

Experiment 3: Pairing Highways and Clocks/Pastries

Method

Participants. Thirty participants (21 females, ages 20-37) took part in Experiment 3A, and an additional 30 participants (23 females, ages 18-39) took part in Experiment 3B. The participants received course credit or payment, and all had normal or corrected-to-normal vision.

Apparatus, Stimuli and Experimental Design. Stimulus display and procedure were identical to these of Experiment 2 with the exception of the type of stimulus categories used:

In Experiment 3A images of highways were paired with clocks, and in Experiment 3B they were paired with pastries. All images were identical to those used in previous experiments, aside from the pastry set that contained 20 bagel images and an additional 20 doughnut images. These were taken from free stock images (pixabay.com, pngimg.com) in accordance with their terms of use, and were modified as needed. The pastry images were presented in grayscale on a grey background, as were the other image categories (see Appendix for image examples).

Results and Discussion

The mean accuracy level in the change detection task of Experiment 3A was 0.67 (SD = 0.08) and of Experiment 3B was 0.69 (SD = 0.09). Figure 4 presents d' means for the four experimental conditions in each of the experiments. We first describe the results of Experiment 3A (highways vs. clocks). A Category (highways vs. clocks) $\times$ Display-type (single- vs. mixed-category) repeated measures analysis of variance (ANOVA) performed on the d' means revealed a significant main effect of category, $F(1, 29) = 13.12, p = .001, \eta_p^2 = .31, BF_{10} > 100$, due to better memory for highways than for clocks. The main effect of display-type approached significance, $F(1, 29) = 3.91, p = .058, \eta_p^2 = .12, BF_{10} = 0.39$, stemming from a slightly better memory in the mixed-category display than in the single-category display (though inconclusive according to Bayesian analysis). Most importantly, the interaction between the two effects was significant, $F(1, 29) = 10.53, p = .003, \eta_p^2 = .27, BF_{10} = 11.49$, resulting from better memory for highways in the mixed-category display than in the single-category display, $t(29) = 3.55, p = .001$, Cohen's $d = 0.65, BF_{10} = 25.36$, while no significant difference was observed between the two display types for the clocks category, $t(29) = -1.54, p > .1$, Cohen's $d = 0.28, BF_{10} = 0.56$ (inconclusive according to Bayesian norms).

The repeated measures ANOVA performed on the d' means of Experiment 3B (highways vs. pastries) demonstrated once again a significant main effect of category, $F(1, 29) = 25.49, p < .001, \eta_p^2 = .47, BF_{10} > 100$, stemming from better overall memory for pastries than for highways. The main effect of display-type approached significance (though inconclusive according to Bayesian analysis), $F(1, 29) = 4.10, p = .052, \eta_p^2 = 0.12, BF_{10} = 0.73$. The category $\times$ display-type interaction was not statistically significant, $F(1, 29) = 1.82, p > .1, \eta_p^2 = .06, BF_{10} = 0.65$. Follow-up contrasts have revealed, however, an asymmetrical mixed-category effect, as the increased memory for highways reached significance within the mixed- relative to the single-category display, $t(29) = 2.57, p = .016$, Cohen's $d = 0.47, BF_{10} = 3.11$, while there was no hint for a difference among the display types in the pastries category, $t(29) < 1, p > .1$, Cohen's $d = 0.04, BF_{10} = 0.20$.

Figure 4. Mean d' for each condition in (a) Experiment 3A and (b) Experiment 3B.

Error bars show ± 1 SE of the mean across participants.

Taken together, the results of the two experiments portray a somewhat different picture from that drawn in findings previously demonstrating a mixed-effect advantage only for faces (and bodies) (Jiang, Remington, et al., 2016). Here, highway images consistently demonstrated a medium to large mixed-category advantage when paired with clocks (Experiment 3A) or with pastries (Experiment 3B). This was in contrast to the findings of Experiment 2A, in which the highway images demonstrated no effect (or even a mixed-category disadvantage) when paired with faces.

Clearly, the mixed-category advantage was smaller in size with the highway images than the facial ones (as observed in Experiments 1 and 2). And yet, the mere fact that the highway category showed different effects, or effects in different directions, when paired with different types of categories, may suggest that the asymmetric nature of the mixed-category advantage may depend on the specific categorical pairing within each experimental setting. That is, rather than being a category-specific effect that heavily relies on the unique perceptual properties of a category (e.g., the configural structure of faces), the mixed-category

advantage is, at least to some degree, an outcome of the interaction between categories embedded within the specific visual display. We suggest that this interaction stems from the relative attentional distribution pattern among the two paired categories. As faces are laden with social and biological significance, and as they may easily bias and/or engage attentional resources (e.g., Langton et al., 2008; Ro et al., 2007, 2001), it is not surprising that a large mixed-category advantage is observed among such type of stimuli when paired with other non-facial categories. The fact that highway images additionally demonstrated a mixed-category advantage when paired with non-facial categories further lends support for this hypothesis.

Note that we did not have prior hypotheses regarding the attentional distribution among categories, and we can only suggest post-hoc speculations as to why highway images were represented more dominantly than the other non-facial categories. One possible reason for scene attractiveness is the richness of the highway images, depicting whole, complex scenes rather than isolated (floating) objects. The fact that all highway images were outdoor scenes, involving a strong depth dimension (due to the highways viewed from the viewer's viewpoint onward), may have additionally contributed to their attractiveness relative to the object categories tested. Importantly, as mentioned above, these accounts are speculative only.

The next two experiments were aimed to further examine the central resource allocation account and to investigate the pattern of the mixed-category effect when modifying the relative attentional distribution across the different categories.

Experiment 4: Modifying the Asymmetrical Mixed-Category Effect by Prior Expectancies

If the attentional distribution across two categories accounts, at least in part, for the asymmetrical pattern of the mixed-category effect, then manipulating attentional allocation to each of the categories should affect the results pattern within the mixed-category trials. That is, categories towards which attention is covertly drawn may show a mixed-category advantage, while those that suffer from a relative attentional deprivation should show no effect, or perhaps even a disadvantage (relative to the single-category trials). Note that such a result pattern should arise regardless of the specific categories at stake, or on top of the inherent tendency of each of the categories to capture and engage attention.

Here, we manipulated the relative relevance of the categories to task requirements, by changing the proportion of trials on which a specific category was tested within the mixed-category displays. That is, we used an endogenous attentional (i.e., expectancy) manipulation that prioritized one of two categories, in an attempt to create an asymmetric pattern of results, similar to the one obtained in earlier experiments. In Experiment 4A, faces and highway images were used, yet now the probability of being tested on a specific category within the mixed-category trials was greater for highways than for faces. We hypothesized that this manipulation would create an asymmetric allocation of attention in favor of the highway images, presumably reducing or cancelling out the more natural tendency to orient to faces. Similarly, in Experiment 4B highways and clocks were used and attention was biased to the clocks, in an attempt to reverse the pattern of the asymmetric results obtained earlier in Experiment 3.

Method

Participants. Thirty participants (23 females, ages 20-38) took part in Experiment 4A, and an additional 30 participants (22 females, ages 21-38) took part in Experiment 4B. The participants received course credit or payment, and all had normal or corrected-to-normal vision.

Apparatus and Stimuli. The apparatus was identical to that of the previous experiments. The stimuli in Experiment 4A were identical to these in Experiment 2A (faces and highways images), and the stimuli in Experiment 4B were identical to these of Experiment 3A (highways and clocks).

Experimental Design and Procedure. The conditions and the trial sequence were identical to these in previous experiments. Since we modified the a-priori probability of the test on each of the categories within the mixed-category trials, the number of trials in the various conditions was changed. The experiment consisted of 360 trials divided into 10 blocks of 36 trials each, preceded by 20 practice trials. In Experiment 4A 60 trials contained a single-category display of four faces, 60 trials contained a single-category display of four highways, 60 trials contained a mixed-category display with a face as the test item, and the remaining 180 trials contained a mixed-category display with a highway as the test item. Thus, within the mixed-category condition, the test item was likely to be a highway image on 75% of trials, while it was likely to be a face image on only 25% of the trials. Similarly, in Experiment 4B, the test item was likely to be a clock image on 75% of trials, while it was likely to be a highway image on only 25% of the trials (within the mixed-category condition). In order to strengthen the task-relevance manipulation, at the beginning of the experiment participants were explicitly notified of this testing bias in the mixed-category displays (no change in instructions was made for the single-category displays). As in previous experiments, within

each condition, half of the trials contained "same" test displays and half contained "different" test displays. All trial types were randomly intermixed throughout the experiment.

Results and Discussion

The mean accuracy level in the change detection task of Experiment 4A was 0.69 (SD = 0.07) and of Experiment 4B was 0.65 (SD = 0.07). Figure 5 presents d' means for the four experimental conditions in each of the experiments. A Category (faces vs. highways) $\times$ Display-type (single- vs. mixed-category) repeated measures analysis of variance (ANOVA) performed on d' means of Experiment 4A revealed a significant main effect of category, $F(1, 29) = 9.74, p = .004, \eta_p^2 = .25, BF_{10} > 100$, resulting from greater sensitivity for faces than for highways, and a significant main effect of display-type, $F(1, 29) = 15.76, p < .001, \eta_p^2 = .35, BF_{10} = 49.09$, stemming from better memory in the mixed-category display than in the single-category display. Critically, in contrast to the robust interaction found in Experiment 2A, there was no hint for an interaction between the two effects in the present experiment, $F(1, 29) < 1, p > .1, \eta_p^2 = .01, BF_{10} = 0.28$. Indeed, both face *and* highway images revealed a mixed-category advantage, $t(29) = 2.29, p = .03$, Cohen's $d = 0.42, BF_{10} = 1.833$ (inconclusive), $t(29) = 3.79, p < .001$, Cohen's $d = 0.69, BF_{10} = 44.57$, respectively.

The repeated measures ANOVA performed on the d' means of Experiment 4B (highways versus clocks) revealed no main effect of category, $F(1, 29) = 2.36, p > .1, \eta_p^2 = .08, BF_{10} = 0.89$, nor of display-type, $F(1, 29) = 1.64, p > .1, \eta_p^2 = .05, BF_{10} = 0.325$. Importantly, the interaction between the two effects was statistically significant, $F(1, 29) = 13.51, p < .001, \eta_p^2 = .32, BF_{10} = 10.23$. A follow-up analysis of the simple effects revealed that in contrast to the mixed-category advantage seen with the highways in Experiment 3A (highways vs. clocks), a mixed-category *disadvantage* was now observed with the same stimuli, $t(29) = -3.13, p = .004$, Cohen's $d = 0.57, BF_{10} = 9.94$. No mixed-category effect was

observed with the clock category, $t(29) = 1.73$, $p = .10$, Cohen's $d = 0.32$, $BF_{10} = 0.73$. It appears, then, that while the clocks category did not gain from the task-relevance manipulation, memory performance in the highway category was impaired in the mixed- relative to the single-category condition.

Figure 5. Mean d' for each condition in (a) Experiment 4A and (b) Experiment 4B.

Error bars show ± 1 SE of the mean across participants.

To further explore the influence of the task contingency manipulation, a statistical comparison was made between Experiment 2A (faces vs. highways) and Experiment 4A (faces vs. highways with a task contingency manipulation), and between Experiment 3A (highways vs. clocks) and Experiment 4B (highways vs. clocks with a task contingency manipulation). Two additional mixed, 3-way ANOVAs were therefore conducted on the d' means. The first analysis included Category (faces vs. highways) and Display-type (single- vs. mixed-category) as within-subject factors, and Experiment (2A vs. 4A) as a between-subject factor. This analysis revealed a significant main effect of category, $F(1, 58) = 35.51, p < .001, \eta_p^2 = .38, BF_{10} > 100$, resulting from overall greater sensitivity for faces than for highways. A significant main effect of display was additionally obtained, $F(1, 58) = 15.47, p < .001, \eta_p^2 = .21, BF_{10} = 17.23$, and so was its interactions with experiment, $F(1, 58) = 4.17, p = .046, \eta_p^2 = .07, BF_{10} = 0.73$ (though inconclusive according to Bayesian norms), and with category, $F(1, 58) = 9.94, p = .003, \eta_p^2 = .15, BF_{10} = 3.29$. The main effect of experiment and its interaction with category did not reach significance ($F_s < 2.3, p_s > 0.1, BF_{s10} < 0.9$). Of most relevance for the current discussion was the statistically significant triple interaction, $F(1, 58) = 15.58, p < .001, \eta_p^2 = .21, BF_{10} = 26.87$, indicating that the patterns of the mixed-category effects within the two experiments differed from each other following the insertion of the task contingency manipulation.

The 3-way mixed ANOVA conducted on Experiments 3A and 4B revealed a significant main effect of experiment, $F(1, 58) = 4.83, p = .03, \eta_p^2 = .08, BF_{10} = 1.76$ (inconclusive), a category $\times$ experiment interaction, $F(1, 58) = 13.60, p < .001, \eta_p^2 = .19, BF_{10} > 100$, and a display-type $\times$ experiment interaction, $F(1, 58) = 5.00, p = .029, \eta_p^2 = .08, BF_{10} = 0.77$ (again, inconclusive). Again, of most interest was the robust triple interaction, $F(1, 58) = 23.56, p < .001, \eta_p^2 = .29, BF_{10} > 100$, suggesting that the different patterns of the mixed-category effects within the two experiments were, at least in part, a function of the relative

relevance of the specific categories within the display. All other effects did not reach statistical significance (all F s < 2.5, p s > 0.1, BFs_{10} < 0.7).

The findings of these two experiments (4A and 4B) and the comparisons between them and their original versions lend support to the central resource allocation account, according to which the asymmetric pattern of the mixed-category effect stems from an unbalanced attentional allocation across the two categories. Changing the relevance of specific categories to task demands (and thus prioritizing one category over the other) altered the pattern of results within the mixed-category condition of each of the categories, relative to their single-category condition. Indeed, the results obtained in Experiments 4A and 4B did not always produce a “classic” asymmetric pattern, as the one observed, e.g., in Experiments 1 and 2 (faces vs. scenes/highways, respectively). Namely, within the faces vs. highways comparison, the mixed category advantage seen originally with faces (Experiment 2A) was not strictly eliminated (though reduced numerically) in Experiment 4A. The highway category, in contrast, flipped its direction: originally showing a mixed-effect disadvantage, it now exhibited a strong advantage effect. Within the highways vs. clocks comparison, the highway category originally showing a mixed-category advantage (Experiment 3A) now showed a disadvantage effect; the clocks category, though, did not gain from the bias of attention, exhibiting no mixed-effect advantage in Experiment 4B. And yet, the alteration of the overall pattern of the mixed, relative to the single category conditions following the task-relevance (i.e., expectancy) manipulation, suggests that the asymmetry of the mixed-category advantage is largely affected by the specific pairings of the two categories and the relative dominance of their representation in working memory. Hence, it appears that rather than relying on the intrinsic characteristics of a category, this asymmetry largely depends on the overall context in which the categories appear.

Note that the present findings support our understanding of the asymmetric mixed-category effect yet they cannot serve as casual evidence for it. Indeed, it may be the case that faces, which are characterized by a fixed configural structure as well as social/emotional significance, demonstrate a mixed-category advantage due to both large inter-categorical competition (as suggested by Jiang, Remington, et al., 2016) and greater attentional engagement. Thus, demonstrating an attentional modification of the face-highway results pattern does not necessitate that attentional bias is the sole underlying mechanism of the mixed-category advantage. As mentioned earlier, however, the fact that scene stimuli (highways), that previously showed no mixed-category advantage, demonstrated such an advantage when prioritized by task demands (Experiment 4A), and at the same time showed a mixed-category disadvantage when serving as an “unattended” category (Experiment 4B), strongly suggests that attention plays an important role in the overall pattern of the mixed-category effect.

Finally, one may argue that the attentional manipulation used in the present experiment simply caused participants to shift their gaze towards prioritized, and away from non-prioritized stimuli, such that the latter were deprived of *visual* processing resources essential for their encoding. Indeed, despite the appearance of a fixation point at the beginning of each trial, the presentation of stimuli for nearly one second allowed participants to move their eyes and fixate on stimuli from one category, at the expense of processing stimuli from the other category. The differences in memory for the different categories may have therefore stemmed from an imbalance in *perceptual*, rather than in attentional resources (though the two are clearly related). To allow better control over perceptual processing of stimuli in both categories, in the next set of experiments we used a sequential, rather than a simultaneous presentation. Note that Jiang, Remington, et al. (2016) tried to control for some spatial attentional factors potentially affecting stimulus encoding, by presenting stimuli sequentially

(at screen center) rather than simultaneously (in different screen locations). A sequential presentation assures that exposure duration and, consequently, processing time of each individual stimulus is equated within and across category types. In addition, the appearance of all stimuli in the same spatial location (i.e., screen center) potentially controls for the locus of spatial attention. Critically, however, such a sequential presentation might not truly prevent an imbalance in the allocation of more *central* attentional resources required for the processing of the different types of categories. Namely, participants may view items from different categories to an equal amount of time, but still allocate different amounts of resources (that are independent of spatial location) for their encoding and/or subsequent retention. The mere presentation of stimuli in a sequential mode, thus, might not prevent faces from engaging central attention resources and eliciting a beneficial performance in mixed-category, relative to single-category trials.

We therefore ran another set of experiments – the first (Experiment 5A) included a simple sequential presentation, that intended to form a baseline for equal perceptual exposure across categories. Two additional experiments (5B & 5C) replicated the first experiment yet with the addition of a task contingency (i.e., attentional bias) manipulation, similar to the one used in Experiment 4. We asked whether despite the sequential presentation conditions in these experiments, biasing attention to one of two categories would modify the asymmetric mixed-category pattern of results, as a function of the relative dominance of categories in working memory representation.

Experiment 5: Presenting Stimuli Sequentially Rather Than Simultaneously

As mentioned above, the present set of experiments was aimed to rule out the possibility that the asymmetric mixed-category advantage observed in Experiment 4 (as well as in previous experiments) was due to differences in the mere exposure and/or perceptual processing of

items in the different categories. Experiment 5A served as a baseline experiment, in which we ran a sequential presentation version of the original Faces vs. Highways study. Critically, in Experiment 5B attention was biased to the highway category, by using a task contingency manipulation. Finally, Experiment 5C examined whether drawing attention to the facial category would further deepen the asymmetry of the mixed-category effect. Note that due to the emergence of the COVID-19 pandemic, Experiment 5C was run online. To assure that the results of this experiment were reliable and comparable to previous laboratory experiments, an additional *baseline* experiment was run online as well, to which Experiment 5C was compared. This experiment was a replication of Experiment 5A (see Footnote #6 below).

Method

Participants. Thirty participants (24 females, ages 18-35) took part in Experiment 5A, an additional 30 independent participants (19 females, ages 19-37) took part in Experiment 5B, and an additional 30 independent participants (24 females, ages 18-40) were recruited to Experiment 5C conducted online. All participants received course credit or payment, and all had normal or corrected-to-normal vision.

Apparatus and Stimuli. Apparatus and stimuli in the laboratory studies were identical to these of Experiment 2A. As for the online experiment, instructions and stimuli were similar to previous experiments. Naturally, there was limited control over experimental conditions such as stimulus viewing distance, etc.

Experimental Design. Trial sequence was as follows: after a 300 ms fixation display, each of the four images was presented at screen center for 200 ms (with no ISI, the presentation order of stimuli was random). This assured that the overall exposure duration of stimuli equaled 800ms, similar to previous experiments. After a 1000 ms blank retention interval, a test image was presented at screen center until response. Experiment 5A consisted

of 288 trials divided into 9 blocks of 32 trials each, which were preceded by 20 practice trials. Experiment 5B consisted of 432 trials divided into 12 blocks of 36 trials each, which were also preceded by 20 practice trials. Within the experimental trials, 72 trials contained a single-category display of four faces, 72 trials contained a single-category display of four highways, 72 trials contained a mixed-category display with a face as the test item, and the remaining 216 trials contained a mixed-category display with a highway as the test item. Thus, within the mixed-category condition, the test item was likely to be a highway image on 75% of trials, while it was likely to be a face image on only 25% of the trials. Experiment 5C was identical to the Experiment 5B, yet now attention was biased to the face category; that is, the test item was likely to be a face image on 75% of the mixed-category trials. All other experimental parameters were identical to previous experiments.

Results and Discussion

The mean accuracy level in the change detection task was 0.72 (SD = 0.04) in Experiment 5A, 0.74 (SD = 0.03) in Experiment 5B, and 0.72 (SD = 0.07) in Experiment 5C. Figure 6 presents d' means for the four conditions in each of the experiments. We first describe the results of Experiment 5A (baseline sequential version). A Category (faces vs. highways) $\times$ Display-type (single- vs. mixed-category) repeated measures ANOVA performed on the d' means of Experiment 5A revealed a significant main effect of category, $F(1,29) = 49.05$, $p < .001$, $\eta_p^2 = .63$, $BF_{10} > 100$, resulting from a higher sensitivity for faces than for highways. The main effect of display-type also reached significance, $F(1,29) = 33.23$, $p < .001$, $\eta_p^2 = .53$, $BF_{10} > 100$, as memory in the mixed-category display was better than in the single-category display. The interaction between category and display-type was significant, $F(1,29) = 17.78$, $p < .001$, $\eta_p^2 = .38$, $BF_{10} > 100$, and it was clarified by further analyses of the simple effects: memory for faces was much better in the mixed-category

display than in the single-category display, $t(29) = 6.99, p < .001$, Cohen's $d = 1.28$, $BF_{10} > 100$, while there was no hint for such a mixed-category advantage in the highways, $t(29) < 1, p > .1$, Cohen's $d = 0.13$, $BF_{10} = 0.25$. These results fully replicate the results obtained in our original study using a simultaneous stimulus presentation (Experiment 2A). The results also replicate Jiang, Remington, et al.'s asymmetrical mixed-category results pattern when using a sequential presentation mode (though note that these authors contrasted faces with scenes from various subcategories, rather than with one subcategory only).

The repeated measures ANOVA performed on the d' means of Experiment 5B, in which attention was biased towards the highway category, showed a significant main effect of category, $F(1, 29) = 63.86, p < .001, \eta_p^2 = .69, BF_{10} > 100$, again, stemming from better memory for faces than for highways. A main effect of display-type was also observed, $F(1, 29) = 35.08, p < .001, \eta_p^2 = .55, BF_{10} > 100$. In contrast to the previous experiment, however, the interaction between category and display-type was largely eliminated and failed to reach statistical significance, $F(1, 29) = 2.09, p > .1, \eta_p^2 = .07, BF_{10} = 0.56$ (inconclusive according to Bayesian norms). That is, a robust mixed-category advantage was observed both for faces, $t(29) = 4.28, p < .001$, Cohen's $d = 0.77$, $BF_{10} > 100$, and for highways, $t(29) = 4.74, p < .001$, Cohen's $d = 0.87$, $BF_{10} > 100$. Again, as in Experiment 4A, the mixed-category advantage was not eliminated among the facial stimuli when inserting the task contingency manipulation, yet a strong mixed-effect *advantage* was now obtained with the highway stimuli.

Figure 6. Mean d' for each condition in (a) Experiment 5A, (b) Experiment 5B, and (c) Experiment 5C. Error bars show ± 1 SE of the mean across participants.

As in Experiment 4, in order to further explore the influence of the attentional (i.e., task contingency) manipulation on the sequential version of the change detection task, a comparison was made between the two sequential experiments: Experiment 5A (faces vs. highways) and Experiment 5B (faces vs. highways with a task contingency manipulation favoring highways). A 3-way ANOVA was conducted on the d' means, including Category (faces vs. highways) and Display-type (single- vs. mixed-category) as within-subject factors, and Experiment (5A vs. 5B) as a between-subject factor. This analysis revealed a significant main effect of category, $F(1, 58) = 108.96, p < .001, \eta_p^2 = .65, BF_{10} > 100$, resulting from overall better memory performance for faces than for highways. A main effect of display-type was significant as well, $F(1, 58) = 68.30, p < .001, \eta_p^2 = .54, BF_{10} > 100$, stemming from greater sensitivity in mixed- than in single-category displays. Also, a marginally significant effect of experiment was found, $F(1, 58) = 3.98, p = .051, \eta_p^2 = .06, BF_{10} = 0.97$, reflecting somewhat better overall performance under the attention manipulation version than the more

basic sequential one. There was a significant interaction between category and display-type, $F(1, 58) = 16.92, p < .001, \eta_p^2 = .23, BF_{10} = 91.88$, but neither of the latter effects interacted with the experimental factor (all F s $< 0.4, p$ s $> 0.5, BF_{10}$ s < 0.3). Again, of most relevance for the current purpose was the small but significant triple interaction, $F(1, 58) = 4.82, p = .032, \eta_p^2 = .08, BF_{10} = 1.47$ (inconclusive according to Bayesian analysis), suggesting that the biasing of attention to the highway category (in Experiment 5B) changed the overall balance between the categories and their representation in working memory, relative to the single-category conditions.

Whereas Experiment 5B examined the allocation of endogenous attention to highways, Experiment 5C examined whether drawing attention to faces would further *increase* the asymmetry of the mixed-category effect. The repeated measures ANOVA performed on the d' means of Experiment 5C showed a significant main effect of category, $F(1, 29) = 51.83, p < .001, \eta_p^2 = .64, BF_{10} > 100$, indicating better memory for faces than for highways. A main effect of display-type was also observed, $F(1, 29) = 14.59, p < .001, \eta_p^2 = .34, BF_{10} = 3.05$, due to improved memory in the mixed-category than the single-category display. The interaction between category and display-type was statistically significant, $F(1, 29) = 14.72, p < 0.001, \eta_p^2 = .34, BF_{10} = 47.44$, reflecting a mixed-category advantage for faces, $t(29) = 5.29, p < .001$, Cohen's $d = 0.97, BF_{10} > 100$, but not for highways, $t(29) = -0.80, p > .1$, Cohen's $d = 0.15, BF_{10} = 0.26$. In order to explore a potential increase in the asymmetry of the mixed-category effect due to the task contingency manipulation, we further compared the results of online Experiment 5C with an online (baseline) version of Experiment 5A⁶. The 3-way ANOVA conducted on these two experiments revealed a

⁶ The online baseline experiment (which was a replication of Experiment 5A) included 30 independent participants (19 females, ages 18-40) that were recruited online, for course credit. The results of this online experiment largely replicated these of (laboratory) Experiment 5A: The mean accuracy level in the change detection task was 0.70 (SD = 0.04), and the repeated measures ANOVA performed on the d' means yielded significant main effects of category, $F(1, 29) = 38.37, p < .001, \eta_p^2 = .57, BF_{10} > 100$, and of display-type, $F(1,$

significant main effect of category, $F(1, 58) = 89.35, p < .001, \eta_p^2 = .61, BF_{10} > 100$, and of display-type, $F(1, 58) = 33.95, p < .001, \eta_p^2 = .37, BF_{10} > 100$, as well as an interaction between these factors, $F(1, 58) = 24.75, p < .001, \eta_p^2 = .30, BF_{10} > 100$. None of the factors, however, interacted with experiment, and there was no triple interaction (all F s $< 2.1, p$ s $> 0.1, BF_{10} < 0.5$) suggesting that biasing attention to the facial category did not enhance the mixed-category asymmetry.

Taken together, the findings of Experiments 5A and 5B support the central resource allocation account and emphasize the key role of attention in the asymmetric mixed-category advantage, even when controlling for spatial location and perceptual exposure factors across the two categories. A robust, natural bias towards faces was seen in Experiment 5A, similar to the one observed when stimuli were presented simultaneously. Critically, however, this bias was yet modified when using a task-relevance (i.e., expectancy) manipulation favoring the highway stimuli (Experiment 5B). Indeed, faces continued to exhibit an advantage (albeit numerically smaller) in the mixed-category display, perhaps due to their ability to capture, or engage attention even when task-demands rendered them as underprivileged. Importantly, this unintentional engagement of faces was now paired with a voluntary attentional orientation towards the highway stimuli, creating a mixed-effect advantage among these stimuli and changing the overall pattern of results. These findings suggest that attention was originally deployed disproportionally to the categories in some *non-spatial* manner that is associated with central resources (see elaborated discussion in the General Discussion). Note that attentional allocation during natural viewing is not an exhaustive process that necessarily consumes all resources; therefore, even if certain types of stimuli involuntarily capture

29) = 19.71, $p < .001, \eta_p^2 = .41, BF_{10} > 100$. In addition, a significant interaction was observed, $F(1, 29) = 10.17, p = .003, \eta_p^2 = .26, BF_{10} = 5.07$, reflecting a mixed-category advantage for faces, $t(29) = 5.21, p < .001$, Cohen's $d = 0.95, BF_{10} > 100$, but not for highways, $t(29) = 1.20, p > .1$, Cohen's $d = 0.22, BF_{10} = 0.38$.

attention (due, e.g., to their biological or social significance), there may still remain sufficient processing capacity to be allocated to other task-relevant stimuli, as seen in Experiment 5B.

As for the attentional manipulation favoring faces over highways (Experiment 5C), no incremental increase in the mixed-category advantage was observed for the facial stimuli, above and beyond the original advantage. We believe that this may reflect a situation of an attentional saturation, resulting from the powerful ability of faces to capture/engage attention under natural conditions. Alternatively, the indifference of faces to the endogenous attentional manipulation may reflect a ceiling effect resulting from non-attentional factors (such as the degree of similarity between stimuli and the test item, the degree of distinctiveness of items, etc.). That is, performance may have reached a point at which any additional processing resources are negligible, given the specific task difficulty conditions.

General Discussion

The findings from three lines of studies revealed that the asymmetric mixed-category advantage, previously ascribed to face-specific characteristics (such as holistic or configural processing, see Jiang, Remington, et al., 2016), is a more general phenomenon that is largely context-dependent. Specifically, we propose that such an asymmetry relies on an imbalance in attentional distribution across two (or more) object categories that compete for processing resources. In our first line of research, we replicated previous findings demonstrating a mixed-category advantage for faces but not for scene (e.g., highway) or object (e.g., clock) basic-level categories. In a second line of studies, using pairings of stimuli containing only non-facial stimuli, we tested the specificity of the effect, and found a mixed-category advantage for highways while no such effect was observed among other object categories. Scenes are not typically characterized by holistic processing therefore, according to the category-specific interference account (Jiang, Remington, et al, 2016) they were not expected

to elicit strong inter-stimulus competition and to display a mixed-category advantage. Our findings suggest that the asymmetric pattern of the mixed-category effect is dependent on the interaction between two given categories, rather than on category-specific attributes, namely, on the *relative* dominance of categories in working memory representation. To further test this hypothesis, in a third line of research we manipulated top-down expectations in the mixed-category condition, essentially biasing attention to one of two categories, at the expense of its counterpart category. Admittedly, the strong natural bias towards facial stimuli was not eliminated when inserting the attentional-bias manipulation, yet the effect was somewhat attenuated. Highways, in contrast, now showed a mixed-category advantage effect. This modification of the results pattern, following the insertion of a task contingency (expectancy) manipulation, suggests once again that attentional distribution may play a dominant role in the asymmetry of the mixed-category effect. A similar modification of the result pattern was observed when using clock (attention-prioritized) vs. highway (attention-deprived) categories. Finally, to assure that the findings above did not result from mere *perceptual* processing deprivation of the non-prioritized category, a sequential presentation experimental version was conducted, in which stimuli were presented at screen center. Whereas a typical asymmetrical pattern of results (in which faces highly benefit from a mixed display) was seen under natural viewing conditions, this pattern was modulated (yielding an advantage for the highways) when inserting a task contingency manipulation that favored the highways. No further advantage for faces was seen when attention was biased towards the facial stimuli in the mixed-category trials.

Taken together, our findings support a central resource allocation account, according to which the asymmetric mixed-category effect likely stems from an attentional bias to one of the two categories. This attentional bias is not necessarily spatial in its nature, and it

presumably affects processing stages subsequent to the initial perceptual encoding phase in working memory (see further discussion below).

Several other lines of evidence suggest that the asymmetry typically characterizing the mixed-category effect is not a category-unique phenomenon. To the extent that faces exhibit large inter-category competition due to their configural nature (according to the category-specific interference account), one would predict maximal stimulus interference when multiple same-category items are presented. That is, performance in the facial single-category condition should be worse than that in the corresponding condition of the highways/clocks. Our findings show, however, that this is not the case, in fact, a reversed pattern is typically observed (i.e., higher memory sensitivity for single category displays of faces than of highways/clocks, see Experiment 2). In contrary, thus, to what could be deduced from Jiang, Remington, et al.'s (2016) account, benefiting from a mixed-category display (due to enhanced within-category competition) is not necessarily associated with a relative reduced performance in the single-category condition. This finding is in accord with previous studies showing a similar pattern of results when using narrowed/homogeneous scene categories (Jiang, Remington, et al., 2016, see pilot study), yet note that several other studies have shown an opposite pattern of results under similar encoding conditions, when using object categories such as cars or watches (Curby & Gauthier, 2007, see short exposure conditions in Experiments 2 and 3; Wong et al., 2008, see Experiments 1 and 4B). The independence of single- and mixed-category memory rates was further observed in the present study, following the insertion of the attentional bias (expectancy) manipulation. While the asymmetric pattern of the mixed-category effect was modulated or even flipped (compared to the same conditions with no attentional manipulation), faces in the single-category display continued to exhibit improved memory performance relative to their corresponding highway counterparts (see Experiment 4A). This orthogonality of the single- and mixed-category conditions thus

supports the view that the mixed-category asymmetry is not a category-specific phenomenon, or the output of within-category competition factors. Rather, it relies on the relative competition *between* categories and on more global attentional distribution factors.

Note that memory performance in a specific category, as reflected in the single-category condition, can be affected by a variety of factors such as the degree of similarity among items within the category. Recently, much research has been devoted to examining the variables affecting the memorability of different stimuli and/or categories (albeit, typically these studies involve long-term memory, rather than working memory, see, e.g., Bainbridge, et al., 2013; Isola, Parikh, et al., 2011; Isola et al., 2014; Isola, Xiao, et al., 2011). A comprehensive study conducted by Bylinskii et al. (2015) found that there are scene categories that are consistently remembered better than other scenes categories. Moreover, within each category, images vary in their memorability rate, based on their specific attributes. An important contribution of the Belinskii et al. (2015) study is the investigation of extrinsic factors that affect the memorability of an image, such as the context in which an image is presented. As expected, the memorability of an image is not a fixed value that depends solely on the image itself, but is also a function of its distinctiveness with respect to its context of presentation (i.e., relatively distinctive images are better remembered). Most important for the present purposes, different categories are affected differently by context. That is, while the average memorability of images increases when presented in the context of other categories, the degree of this increase largely differs among different types of categories (see Figure 5 in Belinskii et al., 2015). While some categories largely benefit from presentation among other-category images, others only show a minimal, or no, memory advantage. We suggest that attentional (and perhaps motivational) factors affect the extent to which a category will profit from a mixed-category display, often creating a situation in

which one or more categories benefit from mixed displays at the expense of other categories, as demonstrated in the present study.

Lastly, there are good reasons to believe that when equating attentional resources across the different categories, *all* categories should benefit from mixed displays, compared to single-category ones. That is, given sufficient visual (and perhaps conceptual) differences among category types, the more distinctive are categories from each other, the better they are encoded and remembered. We believe, however, that since stimuli largely differ in their salience, or their ‘attractiveness’ value, such situations in which attentional distribution is fully balanced between categories are highly rare. Given an unbalanced allocation of processing resources, any potential mixed-category advantage is amplified in the attended category, while it may be cancelled out (or perhaps even inversed) in the unattended category, yielding an asymmetric pattern of results as the ones typically observed. It seems that attention, thus, modulates any potential advantage of mixed- relative to single-category displays.

The locus of attentional effects on working memory representation

What is the precise stage in which attention 'kicks in' and affects working memory performance? One can roughly divide the processes involved in working memory into three stages: encoding – a perceptual stage, maintenance – a post-perceptual retention stage, and the test phase in which a decision is made regarding the contents of memory. As mentioned earlier, presenting stimuli in a sequential manner (Experiment 5) was meant to control for the perceptual processing time of each item within and across category types. In addition, the appearance of all stimuli at screen center potentially restrained the locus of eye fixations and of spatial attention, preventing eye/attentional shifts to one item at the expense of the other. And yet, inserting an endogenous attentional manipulation in which one of two categories was

prioritized at test, managed to modify the asymmetric pattern of results (Experiment 5B), suggesting that any attentional effect likely occurred at a post-perceptual stage and was non-spatial in its nature. Importantly, one may argue that the mere exposure of stimuli in a balanced manner (i.e., controlling for the exposure duration and location of stimuli) does not truly assure that all items were perceived, or encoded, to an equal depth level. Indeed, the fact that one is *observing* a stimulus does not necessitate that the latter is effectively *encoded* for detection or for subsequent memory (e.g., Drew et al., 2013; Neisser & Becklen, 1975). We thus use the term “perception” in its most basic form, as the mere visual exposure to stimuli, while acknowledging the fact that active perception and/or stimulus encoding may involve additional processes that are beyond ‘seeing’ *per-se*. Ruling out an early perceptual stage as the locus of attentional influence strongly leads, thus, to the assumption that any attentional bias process likely occurs at a rather late perceptual, or post-perceptual retention stage⁷. Some support for an attentional involvement at the retention phase comes from studies investigating the *Retro-cue effect*. This effect refers to the improvement in visual working memory tasks (such as change detection) when attention is guided to a to-be-tested item during the retention interval, namely, *after* items have been perceived/encoded and before the test presentation phase (Griffin & Nobre, 2012; Landman et al., 2003; Makovski & Pertzov, 2015; for review, see Souza & Oberauer, 2016). Although several hypotheses regarding the underlying mechanism of the retro-cue effect have been proposed, there is a general consensus that attended representations can be strengthened during the retention interval of working memory (see Souza & Oberauer, 2016). Indeed, while the cue in the retro-cue paradigm is external in

⁷ One may argue that an attentional bias can also occur at the actual test phase. In the simultaneous presentation experiment, for instance, since both categories are presented during test, a dominant category may involuntarily capture attention away from a cued (tested) item belonging to the non-dominant category. In doing so, the ‘attended’ category may impede memory of the ‘unattended’ one. In the sequential presentation study, however, only a single item is presented at test, thus, there are no distractors that may bias attention in favor of one category over the other. Nevertheless, a strong asymmetrical mixed-category effect is obtained, suggesting that attentional involvement occurs earlier in the process.

its nature, we propose that some *internal* mechanism directs attention to specific categories in the mixed-category paradigm, often at the expense of other categories' representations.

Guidance of attention may be mediated by some inherent preference for specific types of stimuli (due, e.g., to their social or biological significance) or due to their current task-relevance. This attentional allocation *can* be spatial in nature, but is not necessarily so, as our sequential presentation experiments show. Clearly, while the proposed account is somewhat speculative, further research is required in order to better understand the exact manner in which internally-driven attentional mechanisms modulate the representation of certain categories or items in working memory.

Finally, the present research has important overall implications for working memory studies using a variety of stimulus types that may differ in their subjective or objective “attractiveness” value. The findings of our study suggest that when participants are briefly presented with multiple-stimulus displays, working memory performance may be dramatically influenced by inter-stimulus competitive relations. Importantly, the risk of eliciting an asymmetric allocation of attention across stimuli is highly increased when using images of real-world objects, that are laden with meaning and/or subjective relevance. Attentional considerations, thus, should be carefully taken into account when selecting stimuli for such studies, or at minimum, should be acknowledged with respect to their potential influence on working memory performance.

References

- Alvarez, G. A., & Cavanagh, P. (2004). The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects. *Psychological Science*, 15(2), 106–111. <https://doi.org/10.1111/j.0963-7214.2004.01502006.x>
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1063–1087. <https://doi.org/10.1037/0278-7393.20.5.1063>
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual Working Memory a Fixed Number of Represents Items Regardless of Complexity. *Psychological Science*, 18(7), 622–628. <https://doi.org/10.1111/j.1467-9280.2007.01949.x>
- Baddeley, A. D. (1966). The Influence of Acoustic and Semantic Similarity on Long-term Memory for Word Sequences. *Quarterly Journal of Experimental Psychology*, 18(4), 302–309. <https://doi.org/10.1080/14640746608400047>
- Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The Instruct Memorability of Face Photographs. *Journal of Experimental Psychology*, 142(4), 1323–1334. <https://doi.org/10.1037/a0033872>
- Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, 116, 165–178. <https://doi.org/10.1016/j.visres.2015.03.005>
- Cheung, O. S., & Gauthier, I. (2010). Selective interference on the holistic processing of faces in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 36(2), 448–461. <https://doi.org/10.1037/a0016471>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J: L. Erlbaum Associates.
- 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*(24), 8955–8960. <https://doi.org/10.1073/pnas.1317860111>
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. <https://doi.org/10.1017/S0140525X01003922>
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. <https://doi.org/10.1177/0963721409359277>
- Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 1–17. <https://doi.org/10.1167/10.4.16>
- Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review*, *14*(4), 620–628. <https://doi.org/10.3758/BF03196811>
- Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*(3), 274–290. <https://doi.org/10.1177/1745691611406920>
- Drew, T., Võ, M. L. H., & Wolfe, J. M. (2013). The Invisible Gorilla Strikes Again: Sustained Inattention Blindness in Expert Observers. *Psychological Science*, *24*(9), 1848–1853. <https://doi.org/10.1177/0956797613479386>
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. <https://doi.org/10.1037/0033-295X.96.3.433>

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, 10(12), 27–27. <https://doi.org/10.1167/10.12.27>
- Griffin, I., & Nobre, A. C. (2012). Orienting attention to locations in mental representations. *Attention, Perception, and Psychophysics*, 74(1), 146–162. <https://doi.org/10.3758/s13414-011-0218-3>
- Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011). Understanding the intrinsic memorability of images (pp. 2429–2437). Granada, Spain.
- Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What Makes a Photograph Memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 1469–1482. <https://doi.org/10.1109/TPAMI.2013.200>
- Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? In *CVPR 2011* (pp. 145–152). IEEE. <https://doi.org/10.1109/CVPR.2011.5995721>
- Jacobs, A. M. (1986). Eye-movement control in visual search: How direct is visual span control? *Perception & Psychophysics*, 39(1), 47–58. <https://doi.org/10.3758/BF03207583>
- Jeffreys, H. (1961). *Theory of probability*. Oxford, U.K.: Oxford University Press.
- Jiang, Y. V., Lee, H. J., Asaad, A., & Remington, R. (2016). Similarity effects in visual working memory. *Psychonomic Bulletin & Review*, 23(2), 476–482. <https://doi.org/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(9), 1399–1411. <https://doi.org/10.1037/xhp0000228>
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, 139(3), 558–578. <https://doi.org/10.1037/a0019165>
- Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, 43(2), 149–164. [https://doi.org/10.1016/S0042-6989\(02\)00402-9](https://doi.org/10.1016/S0042-6989(02)00402-9)
- Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, 107(1), 330–342. <https://doi.org/10.1016/j.cognition.2007.07.012>
- Lin, P.-H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition*, 17(3), 356–372. <https://doi.org/10.1080/13506280701766313>
- Luck, S., & Vogel, J. (1997). The Capacity of Visual Working Memory for Features and Conjunctions. *Nature*, 390(6657), 279–281.
- Makovski, T., & Pertzov, Y. (2015). Attention and memory protection: Interactions between retrospective attention cueing and interference. *Quarterly Journal of Experimental Psychology*, 68(9), 1735–1743. <https://doi.org/10.1080/17470218.2015.1049623>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- Meinhardt-Injac, B., Persike, M., & Berti, S. (2013). Encoding of faces and objects into visual

- working memory. *NeuroReport*, 24(13), 735–740.
<https://doi.org/10.1097/WNR.0b013e328364a417>
- Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic Bulletin and Review*, 20(1), 163–170.
<https://doi.org/10.3758/s13423-012-0313-z>
- Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical Implications of Articulatory Duration, Phonological Similarity, and Phonological Complexity in Verbal Working Memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29(6), 1353–1380. <https://doi.org/10.1037/0278-7393.29.6.1353>
- Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, 7(4), 480–494. [https://doi.org/10.1016/0010-0285\(75\)90019-5](https://doi.org/10.1016/0010-0285(75)90019-5)
- Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception and Psychophysics*, 64(7), 1055–1067.
<https://doi.org/10.3758/BF03194756>
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, 44(4), 369–378. <https://doi.org/10.3758/BF03210419>
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, 16(2), 283–290.
<https://doi.org/10.3758/BF03203943>
- Rensink, R. (2002). Change detection. *Annual Review of Psychology*, 53, 245–277.
<https://doi.org/10.1146/annurev.psych.53.100901.135125>
- Ro, T., Friggel, A., & Lavie, N. (2007). Attentional biases for faces and body parts. *Visual Cognition*, 15(3), 322–348. <https://doi.org/10.1080/13506280600590434>

- Ro, T., Russell, C., & Lavie, N. (2001). Changing Faces: A Detection Advantage in the Flicker Paradigm. *Psychological Science*, 12(1), 94–99. <https://doi.org/10.1111/1467-9280.00317>
- Roper, Z. J. J., Cosman, J. D., & Vecera, S. P. (2013). Perceptual load corresponds with factors known to influence visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 39(5), 1340–1351. <https://doi.org/10.1037/a0031616>
- Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, 105(16), 5975–5979. <https://doi.org/10.1073/pnas.0711295105>
- Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin and Review*, 18(2), 324–330. <https://doi.org/10.3758/s13423-011-0055-3>
- Sato, S., & Kawahara, J. I. (2015). Attentional capture by completely task-irrelevant faces. *Psychological Research*, 79(4), 523–533. <https://doi.org/10.1007/s00426-014-0599-8>
- Shapiro, K. L., & Miller, C. E. (2011). The role of biased competition in visual short-term memory. *Neuropsychologia*, 49(6), 1506–1517. <https://doi.org/10.1016/j.neuropsychologia.2011.02.017>
- Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, and Psychophysics*, 78(7), 1839–1860. <https://doi.org/10.3758/s13414-016-1108-5>
- Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition*, 13(6), 657–665. <https://doi.org/10.1080/13506280500410949>

Thomaz, C. E. (2012). FEI Face Database. Retrieved June 10, 2018, from
<http://fei.edu.br/~cet/facedatabase.html>

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.

<https://doi.org/10.3758/PBR.17.1.59>

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human*

Perception and Performance, 27(1), 92–114. <https://doi.org/10.1037/0096-1523.27.1.92>

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 11. <https://doi.org/10.1167/4.12.11>

Wong, J. H., Peterson, M. S., & Thompson, J. C. (2008). Visual working memory capacity for objects from different categories: A face-specific maintenance effect. *Cognition*, 108(3), 719–731. <https://doi.org/10.1016/j.cognition.2008.06.006>

Appendix

Image examples in the various categories

Scenes (Experiment 1)

Faces (Experiments 1, 2, 4, & 5)

Highways (Experiments 2, 3, 4, & 5)

Clocks (Experiments 2, 3, & 4)

Pastries (Experiment 3)