Vision Research 162 (2019) 8–19

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Visual search for complex objects: Set-size effects for faces, words and cars

Jennifer Hemström[a,b], Andrea Albonico[a], Sarra Djouab[a,c], Jason J.S. Barton[a,*]

[a] Human Vision and Eye Movement Laboratory, Departments of Medicine (Neurology), Ophthalmology and Visual Sciences, Psychology, University of British Columbia, Vancouver, Canada
[b] Faculty of Medicine, Linköping University, Linköping, Sweden
[c] Faculty of Medicine, University of Auvergne, Clermont-Ferrand, France

ARTICLE INFO

Keywords:
Face recognition
Visual word recognition
Object recognition
Response time
Discrimination

ABSTRACT

To compare visual processing for different object types, we developed visual search tests that generated accuracy and response time parameters, including an object set-size effect that indexes perceptual processing load. Our goal was to compare visual search for two expert object types, faces and visual words, as well as a less expert type, cars. We first asked if faces and words showed greater inversion effects in search. Second, we determined whether search with upright stimuli correlated with other perceptual indices. Last we assessed for correlations between tests within a single orientation, and between orientations for a single object type. Object set-size effects were smaller for faces and words than cars. All accuracy and temporal measures showed an inversion effect for faces and words, but not cars. Face-search accuracy measures correlated with accuracy on the Cambridge Face Memory Test and word-search temporal measures correlated with single-word reading times, but car search did not correlate with semantic car knowledge. There were cross-orientation correlations for all object types, as well as cross-object correlations in the inverted orientation, while in the upright orientation face search did not correlate with word or car search. We conclude that object search shows effects of expertise. Compared to cars, words and faces showed smaller object set-size effects, greater inversion effects, and their search results correlated with other indices of perceptual expertise. The correlation analyses provide preliminary evidence supporting contributions from common processes in the case of inverted stimuli, object-specific processes that operate in both orientations, and distinct processing for upright faces.

1. Introduction

Faces and words are two types of objects for which human have extensive visual experience and expertise. Neuroimaging studies show that face and visual word perception activate large cerebral networks (Cohen et al., 2002; Haxby, Hoffman, & Gobbini, 2000) that include the fusiform gyrus, a region that plays an important role in expert visual processing (Weiner & Zilles, 2016). However, a critical difference between visual words and faces is the hemispheric lateralization of their networks, with faces eliciting a stronger response in the right hemisphere and visual words in the left (Cohen et al., 2002; Kanwisher, McDermott, & Chun, 1997). This lateralization is also reflected in neuropsychology. Lesions to the left fusiform region can cause pure alexia (Kleinschmidt & Cohen, 2006; Leff, Spitsyna, Plant, & Wise, 2006), while lesions to the right fusiform gyrus can lead to prosopagnosia, the inability to recognize faces (Barton, 2008; Davies-Thompson, Johnston, Tashakkor, Pancaroglu, & Barton, 2016; Kleinschmidt & Cohen, 2006).

Comparisons between the processing of faces, visual words, and other objects can contribute to our understanding of expert visual perception. This has particularly come to the fore in neuropsychology with the recent development of the many-to-many hypothesis, which proposes that face and visual word processing share and compete for neural resources in both hemispheres of the brain (Behrmann & Plaut, 2013). The hypothesis predicts that alexia is accompanied by mild deficits in face processing, and prosopagnosia by mild reading difficulties (Plaut & Behrmann, 2011). However, a comparison of face and visual word processing in individuals is hampered by the fact that the tests used to diagnose pure alexia and prosopagnosia differ in their measures. While the key diagnostic feature of pure alexia is an elevated word-length effect, the time needed to read a word as a function of the number of letters in the word, (Barton, Hanif, Eklinder Björnström, & Hills, 2014), the diagnosis of prosopagnosia rests primarily on reduced accuracy on face recognition tests, such as the Cambridge Face Memory

* Corresponding author at: Human Vision and Eye Movement (HVEM) Laboratory, Department of Ophthalmology and Visual Sciences – University of British Columbia, 2550 Willow Street, Vancouver, BC V5Z 3N9, Canada.
E-mail address: jasonbarton@shaw.ca (J.J.S. Barton).

https://doi.org/10.1016/j.visres.2019.06.007
Received 18 December 2018; Received in revised form 7 June 2019; Accepted 16 June 2019
Available online 03 July 2019
0042-6989/

Test (Duchaine & Nakayama, 2006) and the Warrington Recognition Memory Test (Warrington, 1984). Indeed, published diagnostic criteria for prosopagnosia concern only the accuracy data (Albonico, Malaspina, & Daini, 2017; Bowles et al., 2009; Duchaine & Nakayama, 2006; Herzmann, Danthiir, Schacht, Sommer, & Wilhelm, 2008). Thus, standard word processing assessments rely mainly on a temporal measure and face perception evaluations on an accuracy measure, rendering comparisons less than ideal.

One of the attractive features of the word-length effect is that it indexes performance as a function of perceptual processing load. Unlike linguistic variables such as part of speech and the frequency of a word's occurrence in the language, the number of letters in a word is a perceptual variable. The more letters, the more perceptual work involved. In healthy subjects the word length effect is minimal, around 5–30 ms/letter, but in alexic subjects it can reach several hundred ms/letter or more (Barton et al., 2014), indicating a reduction in the amount of word (or letter) processing that can be done per unit time by alexic patients.

There has not been a similar test of face perception that indexes performance as a function of the face processing load. In large part this is because there is no straightforward parallel to the word-length effect in faces. While there are words with few letters and words with many letters, there is no variability among almost all faces in the number of features that each possesses.

Rather, to index perceptual load we used a method that varies the number of stimuli processed in a trial, rather than the number of components in a single stimulus (i.e., the number of letters in one word), with response time as the key outcome variable. In this study we showed participants an array of several stimuli, ranging from three to seven, and asked them to indicate if one of the stimuli differed from the others in its identity. The results of each can include not only mean accuracy and mean response time, but also an object set-size effect, expressed as ms/item, which essentially reflects the extra processing time required to do the task when one more object is added to the array. Hence the object set-size effect would reflect the effect of perceptual load for that object type.

This approach is derived from the classic visual search paradigm, in which observers are presented with a display containing a different number of items and have to decide whether a target is present or absent. Changes in accuracy and response times as a function of the set size are the measures of the search performance and have been used to investigate many aspects of perceptual processing (Treisman & Gelade, 1980) and the relationship between vision and attention (Enns & Rensink, 1990; Wolfe, 1994). Initially it was proposed that visual search used either a serial or parallel strategy (Treisman & Gelade, 1980), but subsequent work suggest that this is probably more of a continuum than a dichotomy (Wolfe, 1998, 2003, 2019). A few studies have examined visual search using faces as stimuli. Set-size effects for facial identity or expression are typically larger than those for simpler items (Nothdurft, 1993). Set-size effects have also been used to study a variety of face-related phenomena. For example, the set-size effect is smaller if the subject is searching for their own face (Tong & Nakayama, 1999). Search is faster when the target face is learned as a dynamic moving stimulus than as a static picture (Pilz, Thornton, & Bulthoff, 2006), and set-size effects are smaller when searching for a negative facial expression than for a positive one (Eastwood, Smilek, & Merikle, 2001).

In this report, our goal was to create object search tests for different stimuli, and to evaluate their results for expertise effects, as reflected in inversion effects, processing load differences, and correlations with other, more standard measures of perceptual accuracy or efficiency. We created such object search tests for faces and words, as examples of expert visual processing, as well as for cars, as an example of a different object category. Cars are of particular interest not so much as a category that is representative of basic-level object recognition, but as another category that can allow expression of expertise, though we suspect that this is not often at the high level of face recognition or reading, and is more variable in the population. Like faces and words, cars offer the possibility to populate a recognition space with many exemplars. It is estimated that humans can remember up to 4000 faces (Jenkins, 2017), and average vocabulary estimates range from 20,000 to 35,000 words. The number of car models and their variations over a 50-year period also extends into several thousands. Also, faces and cars are objects for which humans tend to report the most expertise (McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012). Nevertheless, recognition performance is minimally correlated between cars and faces in healthy subjects (McGugin et al., 2012), and there is some evidence from factor analysis and other approaches to suggest that both cars and faces involve separate processing mechanisms that are both distinct from more general object mechanisms operating with novel objects or other object types (Cepulic, Wilhelm, Sommer, & Hildebrandt, 2018; Richler, Wilmer, & Gauthier, 2017).

One key point in object recognition is that cars, faces and words are all visual objects that have an orientation bias, in that they are not vertically symmetric and our experience with them is dominated by the upright orientation. The accumulation of expertise for such objects may thus also show that bias, and possess an orientation-dependency (Diamond & Carey, 1986; Gauthier, Williams, Tarr, & Tanaka, 1998). To determine if such orientation-dependent expertise was reflected in our object set-size effects, we measured the latter for both upright and inverted stimulus presentation. Thus, our first goal was to determine if face and word search would show an inversion effect. Specifically, we hypothesized that face and word set-size effects would show more rapid and accurate performance for upright than inverted stimuli, while car set-size effects would not.

For a second goal, we addressed the question of whether perceptual expertise is reflected in object search with a second approach. We administered other object processing tests that are thought to index expertise for faces, words and cars and hypothesized that accuracy and temporal measures of search would correlate with performance on these tests.

Finally, we used an individual differences approach with correlational analyses to ask how performance on one search test was related to those on the other search tests, as a means of exploring the possible mechanisms that might be operating. Specifically, we asked if there were cross-orientation correlations for the same object type, more than between different object types, as this would imply some degree of orientation-invariant object-specific processing. On the other hand, we also asked if within each orientation there were cross-object correlations. If so, this would suggest that, for that orientation, there may be a significant contribution from common mechanisms that contribute to performance.

2. Method

2.1. Participants

Twenty-four healthy volunteers (12 males, mean age = 42.9 ± 17.5, age range 20–71 years old, all right-handed) participated. All gave written informed consent and were paid 10 dollars/hour for their participation. The protocol was reviewed and approved by the Institutional Review Boards of the University of British Columbia and Vancouver Coastal Health, and carried out in accordance with the principles of the Declaration of Helsinki.

2.2. Stimuli

Three sets of stimuli were created, one each for faces, words and cars, with each set consisting of 84 images. Stimuli for the inverted experiment were created by rotating the same stimuli by 180°. To minimize contributions from low-level image matching, we had images vary not only in the primary dimension relevant to the test, but also in other irrelevant dimensions (Fig. 1). Thus, while the face test required

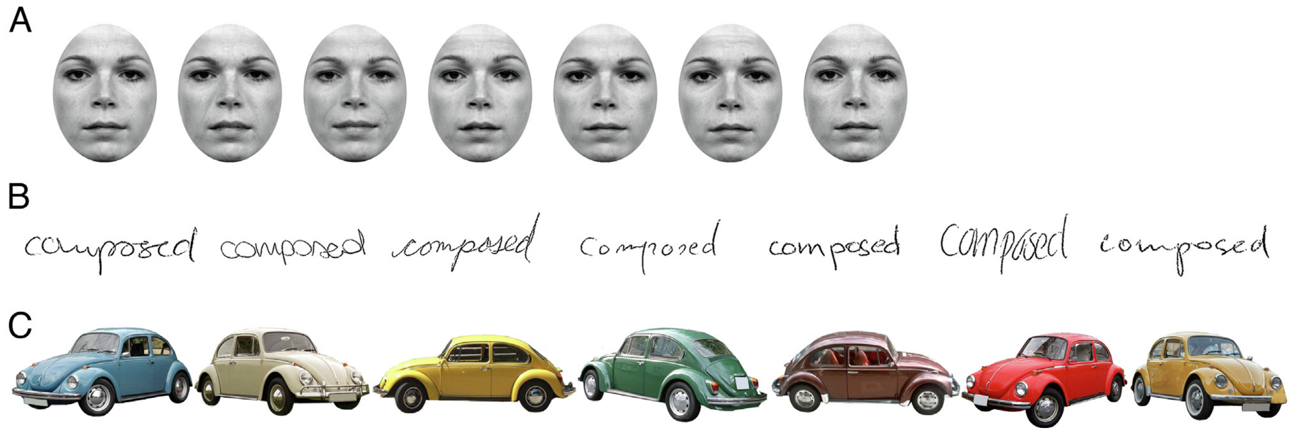

Fig. 1. Examples of stimuli used in the tests: (A) seven different pictures of the same face identity; each picture depict a 80–20% morphing between the neutral face and an emotional face (from left to right: happiness, sadness, surprise, fear, disgust, contempt and anger); (B) seven different handwriting styles for one visual word; (C) seven versions differing in viewpoint and/or colour of one car model.

subjects to report on facial identity, images also varied in facial expression. For the word test, subjects reported on the identity of the word while handwriting varied. For the car test, subjects reported on the car model, while images varied in view and/or colour.

2.2.1. Face stimuli

Facial images from six males and six females were chosen from the Radboud Faces Database (Langner et al., 2010). For each person we obtained seven images, a neutral image and images of six different expressions: happiness, sadness, fear, contempt, disgust and anger. For each person we created more subtle versions of expression by morphing between the neutral face and each of the six emotional faces (80% neutral: 20% expression mix). These six and the neutral face were the seven images used for each person. Each face image was then converted to grey scale and cropped using Adobe Photoshop CC 2014 (www.adobe.com) so that each image was placed in an oval aperture of 195 × 225 pixels, about 5.3 × 6.1° of visual angle. This face size was chosen to ensure that it was in the range in which expert face processing operated (Yang, Shafai, & Oruc, 2014), but small enough to avoid crowding in the displays with larger numbers of faces. This occluded the hair and ears, and textural cues such as spots, scars and freckles were removed. The twelve identities were coupled in 6 pairs (three pairs of female and three pairs of male), matched in age but chosen by pilot work to be different enough that participants could discriminate them with accuracy rates of around 90% (Fig. 2A).

2.2.2. Visual word stimuli

We used twelve eight-letter words in six pairs, chosen to have similar first and last letters and similar global shape, differing from each other in two to four letters (Fig. 2B). The average Kucera-Francis written frequency of these words was 64.4 ± 59.4 occurrences per million visual words (MRC Psycholinguistic Database; http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm). These word pairs were 'starting/standing', 'composed/compound', 'industry/infantry', 'inherent/indirect', 'towering/training', and 'disaster/director'. For each word we collected seven different handwritten versions from seven individuals with different educational and ethnic backgrounds. Each handwritten word was scanned and converted to greyscale using Adobe Photoshop CC 2014 (www.adobe.com). Visual word stimuli had a width of 250 pixels and a height of 117 pixels, or 6.8 × 3.1° of visual angle.

2.2.3. Car stimuli

From the Internet we selected images of twelve car models in six pairs (Lotus Elite/Ferrari 365 GTC/4, Volkswagen Beetle/Porche 911, Volvo 240 GL/Toyota Corolla, Volvo 240 DL wagon/Audi 100 c3 Avant, Smart Car/Fiat 500, Ferrari Dino 246 GT/Jaguar E type, Fig. 2C). These were chosen so that both members of a pair had a similar global shape. For each model, we collected seven pictures differing in viewing angle (side picture, or 2/3 from the front or back) and/or color of the car. Each image was edited using Adobe Photoshop CC 2014 (www.adobe.com) to eliminate the background, the license plate and any identifying lettering or badges denoting the model or the manufacturer. The final car stimuli had a size of about 338 × 145 pixels (9.1 × 3.9° of visual angle).

2.3. Procedure

The tests were controlled by Experiment builder 1.10.1630 (www.sr-research.com) and displayed on a LG monitor (resolution 1920 × 1080 pixels, 52 × 29 cm) at a viewing distance of approximately 57 cm. There were six tests, with each test showing one of the three object types (faces, visual words or cars) in one of two orientations (upright or inverted). Tests were paired according to object type and given in a counterbalanced order across subjects, with half of the participants randomized to start with the upright version for each pair of tests, and the other half to start with the inverted version. Before each of the three test pairs participant were given a test of object processing relevant to that pair. Thus, before doing the face search test they performed the Cambridge Face Memory Test (Duchaine & Nakayama, 2006), before the word search test an assessment of their word-length effect (Albonico & Barton, 2017; Sheldon, Abegg, Sekunova, & Barton, 2012) and before the car search test a car expertise test, based on a semantic car knowledge questionnaire (Barton, Hanif, & Ashraf, 2009).

2.4. Face, word and car search tests

Participants performed a same/different task. The procedure of each test and its trials were the same for all six tests. All stimuli were presented against a white background and a keyboard was used to collect participants' responses. Each trial started with a 250 ms blank screen that was followed by the target screen. The target screen showed an array of a number of stimuli of a single type (i.e., all face, all words, or all cars), ranging from three to seven (Fig. 3). On 'same' (target absent) trials, all the stimuli were examples of the same person, word or car, though each image had a different expression, handwriting, or viewpoint/colour, respectively. On 'different' (target present) trials one of the stimuli was an image of the other member of the pair. Participants were asked to press the key 'S' on the keyboard if all the stimuli were of the same person, word or car, and the key 'D' if one image differed from the rest. The target screen remained on the screen until participant's

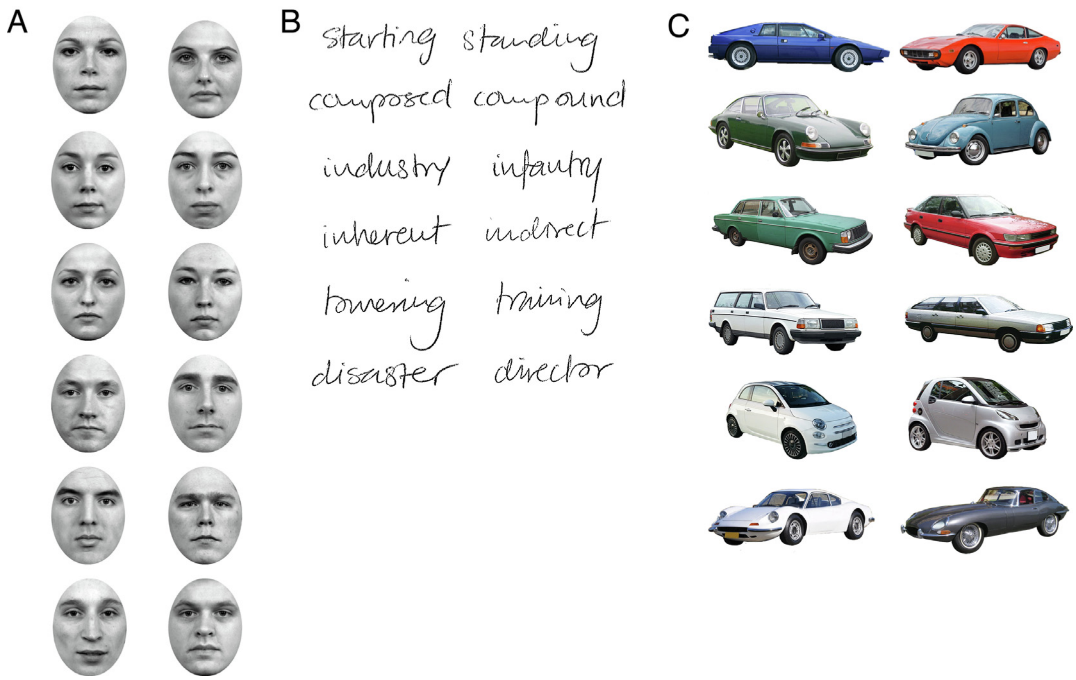

Fig. 2. Examples of the six objects pairs used in the search tasks: (A) face pairs, (B) cars pairs and (C) visual words pairs.

response.

Each test began with a practice phase that first showed a power point presentation that introduced the test and familiarized the participants with the stimuli. After this presentation, participants performed 12 practice trials that showed at least two of each set-size and used some of the same stimuli as the experiment. They received auditory feedback after each response.

A total of 270 experimental trials were completed in each test: 90 trials in which all the stimuli depicted the same face identity, visual word or car, and 180 trials in which one of the stimuli on the screen was different than the others. The 270 trials were divided in three blocks of 90 to give participants breaks during each test. Trials were equally distributed among the number of stimuli in the array (3, 4, 5, 6 or 7 stimuli), giving 54 trials per number. The presentation of the 84 stimuli (faces, visual words or cars) was balanced so that each stimulus appeared the same number of times in each test. Accuracy and response time (defined as the time between the appearance of the stimuli and the participant's response) were recorded.

2.5. Probes of face, word and car expertise

The Cambridge Face Memory Test (Duchaine & Nakayama, 2006) has been used to classify subjects along a spectrum of facial expertise, ranging from developmental prosopagnosia to super-recognizers

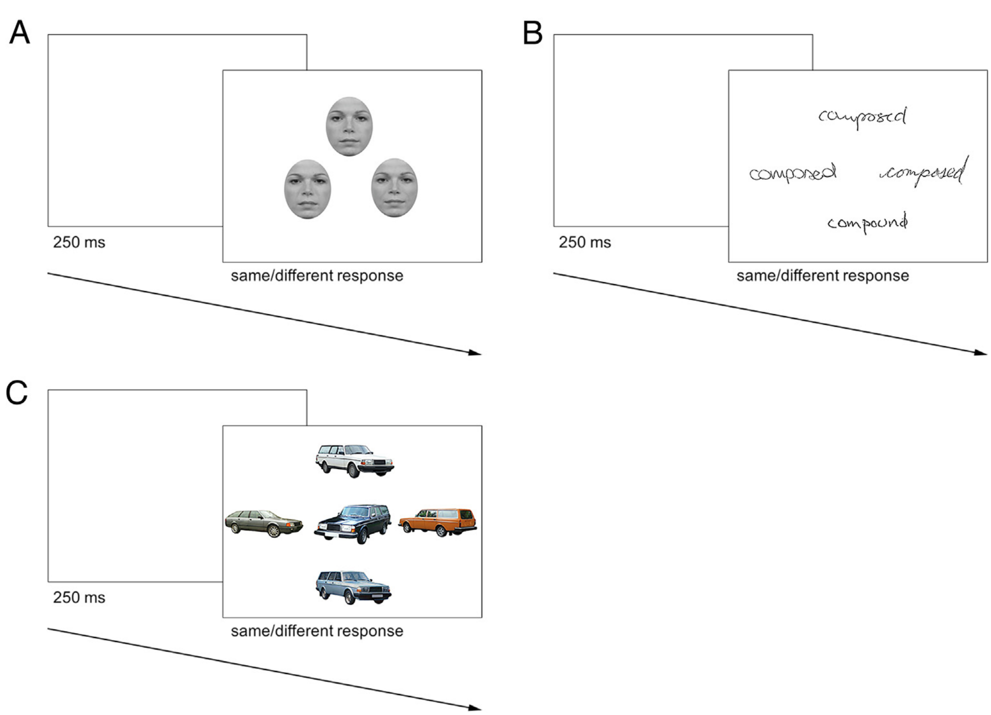

Fig. 3. Illustrations of a typical trial in the visual search tests. Each trial started with a blank screen for 250 ms, followed by a target screen that was displayed until the subject made a response whether all the stimuli were of the same individual, word or car, via key press: (A) a face-search trial with three stimuli on the screen, (B) a word-search trial with four stimuli, and (C) a car-search trial with five stimuli. Note that the sizes of the stimuli in the figure do not reflect the actual sizes used.

(Russell, Chatterjee, & Nakayama, 2012). Participants study six greyscale male target faces during an encoding phase. In a retrieval phase, trials show a target face and two distractors faces, and the subject indicates which was target face. The test includes stages that increase difficulty by adding noise and varying viewpoint. The total score is out of 72.

Naming times and word-length effects are related variables in reading performance (Barton et al., 2014) and studies show that the word-length effect diminishes as children gain in reading proficiency (Aghababian & Nazir, 2000; Zoccolotti et al., 2005). As adults acquire perceptual proficiency with reading upside down, their word-length effect for inverted words decreases (Ahlen, Hills, Hanif, Rubino, & Barton, 2014). Hence temporal measures of reading can be taken as indices of visual word expertise, and loss of such expertise characterizes acquired alexia (Barton et al., 2014). Our word-reading test (Sheldon et al., 2012) presents 60 words, 15 each for words of 3, 5, 7 or 9 letters in length. Participants read each word aloud as quickly as possible and an Andrea NC-8 microphone (http://www.andreaelectronics.com) recorded the participant's vocal response. The time between the appearance of the word and the onset of their vocal response is the naming time. Outcome variables are the mean naming time, and the word-length effect, which is the slope of the regression of response times as a function of the number of letters in the word.

To index car expertise in prior studies, we developed a questionnaire that probed a subject's semantic knowledge about cars (Barton et al., 2009). The car expertise test in this study is a short version of that instrument. In the original questionnaire participants were given a list of 457 models of car and asked to write the name of the manufacturer of each model. For the present study 100 out of the 457 original questions were selected so that those 100 questions gave a correlation of 0.92 with the total verbal expertise score based on the entire sample of 457. This was done on a cohort of 17 healthy participants (12 male, mean age 43.6 ± 16.9 years, range 25–65). A similar use of semantic knowledge to quantify expertise with an object category has been reported by others (Van Gulick, McGugin, & Gauthier, 2016). Results on this verbal semantic test can be considered a proxy index of perceptual car recognition, since semantic scores and perceptual scores have shown a correlation of 0.94 (Barton, Albonico, Susilo, Duchaine, & Corrow, 2019).

2.6. Data analyses

For each of the six blocks we analyzed four outcome variables, two reflecting accuracy and two being temporal measures. For accuracy we calculated the mean accuracy for both same (target absent) and different (target present) trials together. Next we used the hits (participant responded 'different' on target-present trials) and false alarms (participants responded 'different' on target-absent trials) to calculate *d'*, a criterion-free index of discrimination sensitivity (Macmillan & Creelman, 1991). For temporal measures we used only correct trials, and first calculated the mean response time for each block in each subject. Next, we calculated the object set-size effect, which was the slope of the linear regression between the response time and the number of objects in the array.

For each of the four variables we submitted the data to a two-way repeated-measures ANOVA, with object type (face, visual word and car) and orientation (upright and inverted) as within-subject factors. Significant differences were further explored by Bonferroni post-hoc multiple comparisons (corrected *p-values* are reported), in particular comparing upright and inverted performance for each stimulus type. Effect sizes were measured by computing the Partial Eta Squared (η).

Since a broader goal is to create tests that evaluate face, object and visual word processing in a similar manner, we assessed the reliability of each object search test. Internal consistency was measured by Cronbach's alpha (Cronbach, 1951) for both accuracy and response time data, and by computing within-subjects split-half (odd vs. even trials) correlations with Spearman–Brown corrections for each of the four outcome variables from each task.

We also performed two individual differences analyses looking at correlations between the results of different tests. Our first approach investigated whether any of the four outcome variables of the three upright tests correlated with results on other probes of the proficiency of subjects with these object types. Thus, we examined whether performance on the face-search test correlated with accuracy on the Cambridge Face Memory Test. Results on the word-search test were correlated with the results of the reading test, namely the mean naming time and the word-length effect. For cars, we correlated results on the car-search test with the measure of semantic car knowledge obtained from the car expertise test.

Our second approach was to examine the relationships between the results on these six different object search tests. We asked whether performance within a single orientation was correlated among the different object types, which might suggest an element of common processing that generalizes across object categories in that orientation. We also assessed whether performance for a specific object type was correlated between upright and inverted presentations, which would suggest some orientation-invariant processing mechanism for that object.

3. Results

3.1. Upright performance

To first determine if task difficulty was approximately similar across the three search tasks in their typical orientation, we analyzed the data from the tests with upright stimuli with a one-way repeated-measures ANOVA, with object type (face, visual word and car) as a within-subjects factor.

As desired, mean accuracy was above 90% correct for all three object types. Nevertheless, for mean accuracy the effect of object type was significant ($F_{(2,46)} = 4.36$, $p = .018$, $\eta = 0.159$): being slightly better for visual words ($M = 0.957 \pm 0.03$) than for cars (0.903 ± 0.09), while the mean accuracy for faces (0.927 ± 0.08) did not differ from that for either cars or words. The *d'* results also showed a similar effect of object type ($F_{(2,46)} = 3.55$, $p = .037$, $\eta = 0.134$), due to better *d'* for visual words (3.67 ± 0.67) than for cars (3.14 ± 1.06), with faces (3.48 ± 0.92) not differing from the other two types.

For mean response times, there was a trend to an effect of object type ($F_{(1.38,31.74)} = 3.47$, $p = .059$, $\eta = 0.131$),[1] due to a trend for response times for cars (4283 ± 2332 ms) to be slower than response times for words (3264 ± 1103 ms, $p = .063$), while the response times for faces (3324 ± 1434 ms) did not differ from the other two. The object set-size effect did show an effect of object type ($F_{(2,46)} = 4.58$, $p = .015$, $\eta = 0.166$), due to a larger object set-size effect for cars (449 ± 267 ms/item) than for either faces (287 ± 195 ms/item, $p = .027$) or visual words (359 ± 190 ms/item, $p = .039$).

These results indicate that our tests were reasonably well calibrated to give similar accuracy for all three object types, but that the test with cars was slightly more difficult than those for the more expert object categories of words or faces, which did not differ from each other.

3.2. Accuracy and d' (Fig. 4)

Mean accuracy was similar for faces, words and cars, as there was no main effect of object type ($F_{(2,46)} = 0.53$, $p = .592$, $=0.023$). The main effect of orientation was significant ($F_{(1,23)} = 53.30$, $p < .001$, $=0.699$) due to better accuracy for upright stimuli. There was an

[1] Since the Mauchy's test revealed a violation of the sphericity assumption, the degrees of freedom reported are corrected according to the Greenhouse-Geisser method.

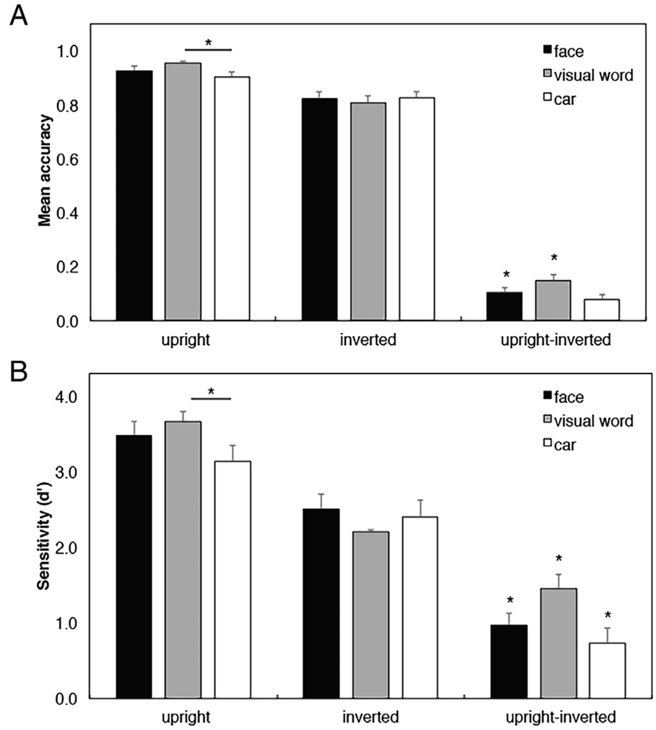

Fig. 4. Accuracy outcome variables for the three object types. (A) Mean accuracy data and (B) (*d′*) data. Each graph shows upright performance, inverted performance, and, on the right, the inversion effect, which is the subtraction between upright and inverted scores. Error bars indicate one standard error. * indicates significance at $p < 0.05$.

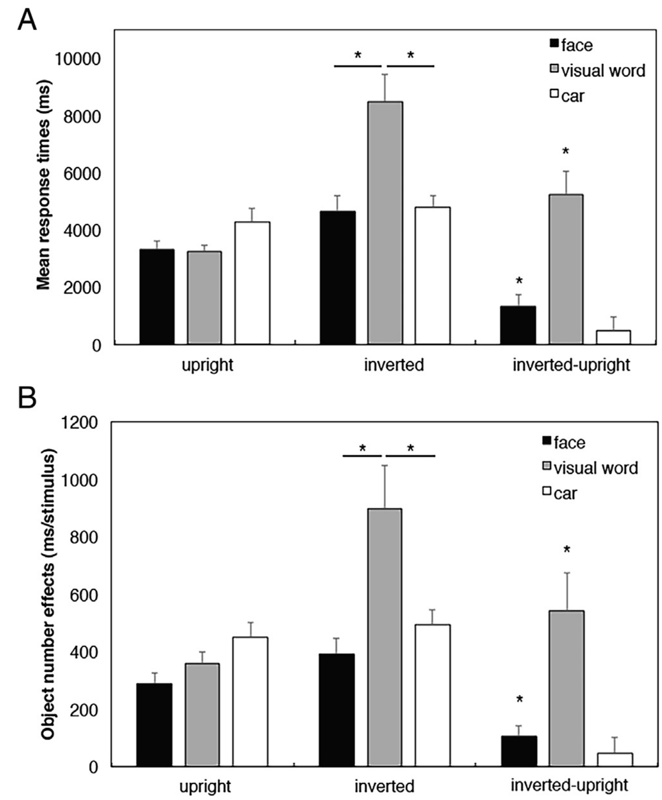

Fig. 5. Temporal outcome variables for the three object types. (A) Mean response times and (B) object set-size effects. Each graph shows upright performance, inverted performance, and, on the right, the inversion effect, which is the subtraction between upright and inverted times. Error bars indicate one standard error. * indicates significance at $p < 0.05$.

interaction between object type and orientation ($F_{(2,46)} = 4.87$, $p = .012$, $\eta = 0.175$): There were differences between upright and inverted performance for faces and words but not for cars.

Likewise, the data for *d'* showed no main effect of object type ($F_{(2,46)} = 0.77$, $p = .468$, $\eta = 0.032$) but a main effect of orientation ($F_{(1,23)} = 79.13$, $p < .001$, $\eta = 0.775$), due to superior discrimination for upright stimuli, and an interaction between object type and orientation ($F_{(2,46)} = 5.61$, $p = .007$, $\eta = 0.196$). In the upright orientation *d'* for cars was lower than that for words, but did not differ from that for faces. There was no difference among the three object types in the inverted orientation. All three object types showed a significant inversion effect (all ps < 0.001).

3.3. Mean response times and object set-size effects (Fig. 5)

For mean response times there was a main effect of object type ($F_{(2,46)} = 9.77$, $p < .001$, $\eta = 0.298$) due to slower responses to words than to faces or cars. There was a main effect of orientation ($F_{(1,23)} = 25.06$, $p < .001$, $\eta = 0.521$) with faster responses to upright stimuli. There was an interaction between object type and orientation ($F_{(2,46)} = 29.92$, $p < .001$, $\eta = 0.565$). Response times for cars did not significantly differ between the upright and inverted orientation ($p = .290$), but those for words and cars did, with faster responses to upright stimuli (both ps < 0.01).

For object set-size effects there was a main effect of object type ($F_{(2,46)} = 9.10$, $p < .001$, $\eta = 0.284$), with the word set-size effect being larger than the face set-size effect but not the car set-size effect. There was a main effect of orientation ($F_{(1,23)} = 13.44$, $p = .001$, $\eta = 0.369$), with more prolonged object set-size effects for inverted stimuli. The interaction between object type and orientation was also significant ($F_{(2,46)} = 13.66$, $p < .001$, $\eta = 0.373$). Object set-size effects did not differ between object types in the upright orientation, but the word set-size effect was larger than the face- and car set-size effects with inverted stimuli. Object set-size effects for cars did not significantly differ between the upright and inverted orientation ($p = .427$), but those for words and faces did (both ps < 0.05), with larger object set-size effects when words and faces were inverted.

3.4. Reliability analyses (Table 1)

Cronbach's alpha was excellent for the accuracy and response time measures of all tests, being above 0.80 in all. However, since a large number of test items (i.e., trials in the present case) can artificially inflate alpha (Cortina, 1993), such values should be interpreted carefully. Split-half correlations with Spearman–Brown corrections confirmed substantial internal consistency for mean accuracy, d', mean response times and set-size effects across all six object-search tests, with all correlations were above 0.70 and significant at $p < .01$. Split-half correlations for the set-size effects were somewhat reduced compared to the other measures. This might reflect the fact that set-size effects combine data from different trials into a single index, which might attenuate the correlation coefficients (Overall & Woodward, 1975).

3.5. Individual differences 1: correlations with tests of proficiency for faces, visual words and cars

Average score for the twenty-four participants on the Cambridge Face Memory Test was 56.2 ± 11.3. These scores correlated with the mean accuracy ($r = 0.609$, $p = .002$) and *d'* ($r = 0.540$, $p = .006$) on the face search test, but with only a trend for the mean response time ($r = -0.39$, $p = .059$) and not with the face set-size effect ($r = -0.28$, $p = .192$).

In the word reading test participants had a mean naming time of

Table 1
Cronbach's alpha values and split-half correlations with Spearman–Brown corrections for all the six object search tests.

	Cronbach's alfa		Split-half reliability			
	Accuracy	Response time	Accuracy	d'	Response times	Set-size effect
Upright						
Faces	0.964	0.993	0.975	0.944	0.993	0.837
Visual Words	0.829	0.985	0.801	0.747	0.985	0.853
Cars	0.978	0.993	0.973	0.957	0.989	0.826
Inverted						
Faces	0.974	0.993	0.975	0.929	0.991	0.705
Visual Words	0.970	0.989	0.957	0.949	0.992	0.882
Cars	0.966	0.988	0.975	0.944	0.983	0.751

673 ± 153 ms and average word-length effect of 6.6 ± 7 ms/letter. Mean naming time correlated with the mean response time ($r = 0.772$, $p < .001$) and word set-size effect ($r = 0.604$, $p = .002$) on the word search test, but not with the mean accuracy ($r = 0.29$, $p = .156$) or d' ($r = 0.30$, $p = .154$). No significant correlations were found with the word-length effect in reading.

In the car expertise test, participants had an average verbal semantic score of 14.74 ± 14.81. There was no correlation between this semantic score and any of the four outcome variables of the car search test.[2]

3.6. Individual differences 2: correlations between face, word and car search results (Table 2);

Correlations between stimuli in different orientations can suggest orientation-invariant object processing mechanisms. For all four measures, the results for all object types showed a significant correlation between upright and inverted orientations for the same object type (dark boxes, Table 2), but weaker or non-existent correlations between one object type in one orientation and a different object type in the other orientation. In aggregate, the cross-orientation correlations between the same object types were greater than those between different object types (Mann-Whitney $U = 23$, $z = 4.06$, $p < 0.0001$). To show this we performed a *meta*-analysis of correlation coefficients using Fisher Z as effect size (Hedges & Olkin, 1985), and plotted the results as a Forest plot (Fig. 6A).

Correlations between object types for a single orientation may suggest common mechanisms that operate on different stimuli in that orientation (outlined boxes, Table 2). For inverted stimuli, many significant correlations were observed: indeed, correlations between all stimuli were significant for accuracy and response time. In contrast, for upright stimuli, there was no correlation for any measure between faces and either cars or words, though upright words and upright cars showed a consistent correlation. Overall, there was a trend for correlations for inverted stimuli to be larger than those for upright stimuli (Mann-Whitney $U = 40$, $z = 1.85$, $p = 0.064$, Fig. 6B).

3.7. Target present and target absent trials

As stated in methods, our analyses above report results in which mean accuracy, d', mean response time, and object set-size effect are computed combining the data from both target-present and target-absent trials. Analysis of the results for target-present and target-absent trials separately produced results generally similar to the analysis combining the two. However, a few points deserve comment (Table 3).

First, errors include both misses and false alarms. This is consistent with low-threshold theories of search, such as signal detection theory (Eckstein, Thomas, Palmer, & Shimozaki, 2000), in which noisy representations of targets and distractors fall along a continuum with the potential for overlap (Palmer, Verghese, & Pavel, 2000). In contrast, false alarms do not occur in high-threshold theory. Set-size effects such as we found do not occur in high-threshold theory but are characteristic of low-threshold theories (Palmer et al., 2000).

Second, the set-size effect is larger for target-absent trials than target-present trials for all stimuli except inverted faces. Our set-size effects are much larger than the typical 20–60 ms/item reported for search with simpler visual stimuli and most likely incorporate the time to execute searching saccades (Wolfe, 1998). Nevertheless, the ratio of set-size effects of target-absent over target-present trials is close to 2 for a number of conditions, namely upright faces and upright or inverted words (Table 3). A ratio of 2 is predicted by termination rules in the standard serial self-terminating search model, in which exhaustive search is needed to reach a correct decision on target-absent trials, while search on target-present trials can be terminated by discovery of the target, which will occur on average about halfway through scanning (Cousineau & Shiffrin, 2004; Townsend & Wenger, 2004; Wolfe, 1998). Lower ratios of around 1.5, similar to those we found for upright or inverted cars, have been seen in searches with premature termination on target-absent trials, possible due to a contribution of guidance from parallel pre-attentive mechanisms (Cousineau & Shiffrin, 2004). However, it is not clear what effects would account for a ratio of 1, as we found for inverted faces.

4. Discussion

The preliminary practical aim of this study was to develop an assessment of face, word and car perception that had similar accuracy and temporal variables as primary outcomes, and indexed the processing load involved in perception of these stimuli. For upright stimuli, our tests were successfully calibrated to give similar mean accuracy of 90 to 96% for all three stimuli. Hence our stimulus sets are reasonably similar in level of difficulty and provide similar numbers of correct trials for response time analysis. Interestingly, with these upright stimuli the object set-size effect was greater for cars than for the more expert categories of words or faces. This suggests that one of the effects of expertise is more efficient handling of perceptual processing load.

Our main goal was to examine whether our outcome variables showed inversion effects. We found consistent effects of orientation in all variables, including the two measures of accuracy, mean accuracy and d', as well as the two temporal measures, mean response time and object set-size effect. All four measures showed that, for words and faces, performance was worse with inverted stimuli, while cars did not show an effect of orientation. This would be consistent with the long-standing view that the expertise gained through long experience with objects biased towards a typical orientation results in an 'inversion effect', in which perceptual performance for the typical orientation exceeds that for atypical orientations. Indeed, many studies showed that

[2] We also computed cross-category correlations between expertise measures and object search performance. However, no cross-category correlations were found between expertise measures and any of the accuracy or temporal outcome variables in the object search tests.

Table 2
Correlations between face, visual word and car search results.

ACCURACY OUTCOME VARIABLES

accuracy		upright face	upright car	upright word	inverted face	inverted car	inverted word
upright	face		0.179	0.148	**0.712******	0.207	0.358
upright	car			**0.626***	**0.468***	**0.633*****	**0.472***
upright	word				0.315	0.309	**0.656******
inverted	face					**0.538****	**0.614******
inverted	car						**0.512****
inverted	word						

d'		upright face	upright car	upright word	inverted face	inverted car	inverted word
upright	face		0.323	0.236	**0.709******	0.122	0.236
upright	car			**0.710***	**0.468***	**0.632*****	**0.611*****
upright	word				**0.410***	0.390	**0.689******
inverted	face					0.345	**0.451***
inverted	car						**0.582*****
inverted	word						

TEMPORAL OUTCOME VARIABLES

response time		upright face	upright car	upright word	inverted face	inverted car	inverted word
upright	face		0.022	0.353	**0.711******	0.385	0.327
upright	car			**0.504***	0.113	**0.460***	0.231
upright	word				**0.519****	**0.659*****	**0.576*****
inverted	face					**0.501***	**0.768******
inverted	car						**0.639******
inverted	word						

set-size effect		upright face	upright car	upright word	inverted face	inverted car	inverted word
upright	face		-0.033	0.280	**0.725******	0.090	0.391
upright	car			**0.656***	0.059	**0.429***	0.118
upright	word				**0.485***	0.292	**0.473***
inverted	face					0.226	**0.637******
inverted	car						**0.526****
inverted	word						

Dark boxes indicate correlations between upright and inverted orientations for the same object type.
Outlined boxes indicate correlations between object types for a single orientation.
Bold indicates significant values.
* Indicates $p < .05$.
** Indicates $p < .01$.
*** Indicated $p < .005$.
**** Indicates $p < .001$.

that the inversion effect is typically greater for objects of expertise than for other non-expert objects (Diamond & Carey, 1986; Gauthier et al., 1998; Valentine, 1988; Yin, 1969). The fact that we find an inversion effect for face and words but not for cars suggests that the object search paradigm does draw upon a subject's perceptual expertise. In particular, this pattern of orientation results in the object set-size effect indicates that perceptual expertise is reflected in the effect of perceptual load on processing times.

To further support a role of expertise in our findings, the results also showed correlations between the object search results and other measures that probe perceptual expertise. Accuracy on the Cambridge Face Memory Test correlated with accuracy (but not temporal indices) on the face search test, while the naming time on the word reading test correlated with temporal indices (but not accuracy) on the word search test. We draw two conclusions from these results. First, this supports the conclusion from our inversion effect data that face and word expertise effects are captured in the performance of our object search tests. Second, they reinforce the concern that accuracy and temporal parameters may not correlate well with each other, and underscores one key motivation behind this study, the need to find parallel outcome variables for face and word processing that can support meaningful comparisons.

Two negative results deserve some brief comment. First is the lack of correlation of word search results with the word-length effect. This likely reflects the modest size and relative homogeneity in that index in healthy subjects. However, there is a systematic relationship between

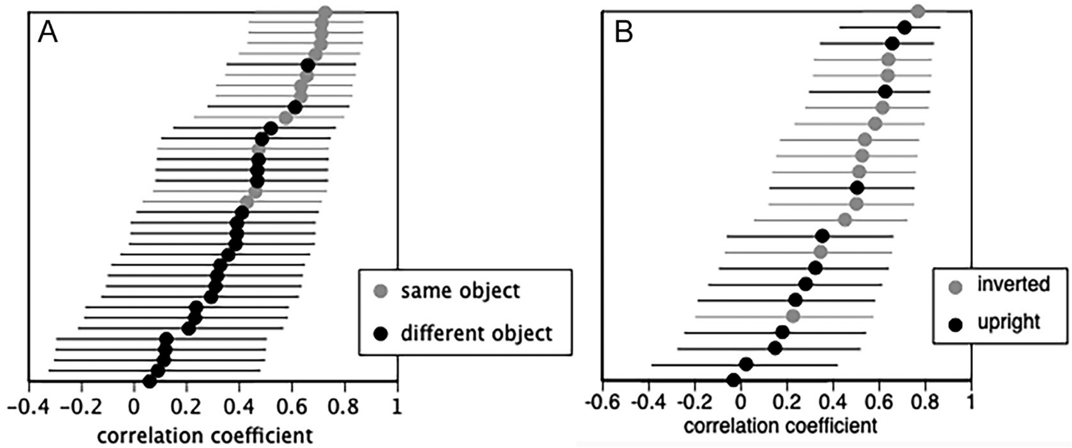

Fig. 6. Forest plots of meta-analysis of correlation coefficients listed in Table 2. A. Cross-orientation comparison of correlations for the same object type (grey symbols) versus correlations between different object types (black symbols). B. Same-orientation correlations between different object types, comparing upright (black symbols) and inverted (grey symbols) presentations. Data for all four different outcome variables (accuracy, d', response time, set-size effect) are included in both plots. R-values are plotted, with bars spanning their 95% confidence interval, sorted from lowest at bottom to highest at top.

Table 3
Results for the target-present and target-absent trials.

	FACES				WORDS				CARS			
	Accuracy		Set-size effect		Accuracy		Set-size effect		Accuracy		Set-size effect	
	Mean	*s.d.*	Mean	s.d.	Mean	*s.d.*	Mean	s.d.	Mean	s.d.	Mean	s.d.
UPRIGHT												
Target present	**0.92**	0.11	**232**	204	**0.96**	0.03	**286**	204	**0.90**	0.08	**379**	231
Target absent	**0.95**	0.05	**397**	268	**0.96**	0.04	**558**	269	**0.91**	0.13	**601**	406
p (1)			0.003				< 0.0001				0.001	
Ratio (absent/present)			1.95				2.13				1.56	
p (2)			0.536				< 0.001				< 0.0001	
INVERTED												
Target present	**0.81**	0.17	**405**	316	**0.81**	0.14	**685**	606	**0.81**	0.13	**417**	239
Target absent	**0.88**	0.14	**412**	353	**0.83**	0.16	**1270**	1089	**0.87**	0.13	**637**	391
p (1)			*0.923*				*0.001*				*0.004*	
Ratio (absent/present)			0.95				1.94				1.60	
p (2)			*< 0.0001*				*0.351*				*< 0.0001*	

p (1) probability that the subtraction (target absent RT) − (target present RT) differs from 0.
p (2) probability that ln(target absent RT/target present RT) differs from natural logarithm of 2 (ln(2)).

word-length effect and mean naming time across a wide range of healthy and impaired subjects (Barton et al., 2014), and it may be that mean naming time has better signal-noise ratio in healthy subjects, given the very small size of word-length effects in healthy literate subjects. To broaden the range, a study in developmental dyslexia, acquired alexia or even children learning to read may clarify the relationship between word search performance and word-length effects.

Second is the lack of correlation between the results of the car search test and car semantic knowledge. However, the latter is a rather indirect probe of car expertise that, unlike the Cambridge Face Memory Test and the reading test, does not explicitly probe perceptual function. Furthermore, most of our subjects were not car experts by this test, with five even scoring zero and only seven in the expert range. Finally, our sample is only of a modest size. For all of these reasons, failure to find a correlation must be treated cautiously and cannot be taken as definitive evidence of a lack of a relationship between car search and car expertise.

Our last individual differences analysis, of the correlations between the search tests, also yielded interesting preliminary results. First, there were correlations between upright and inverted accuracy measures within each object type, more so than between object types, suggesting a contribution from object-specific, orientation-invariant processes. Thus face-specific processing skills contribute strongly to face search accuracy regardless of orientation, and the same is true for words and cars. Such cross-orientation correlations could argue against a complete switch between very different processing mechanisms for upright and inverted stimuli, particularly for faces as suggested by the dual-mode hypothesis (Bartlett & Searcy, 1993), but could support a quantitative difference, a reduction in efficient use of an expert mechanism when confronted by stimuli in atypical orientations (Albonico, Furubacke, Barton, & Oruc, 2018; Sekuler, Gaspar, Gold, & Bennett, 2004).

Second, performance with inverted stimuli frequently correlated across object types, for both accuracy and temporal measures. This suggests a substantial contribution from shared mechanisms when inverted stimuli are processed. While a shared mechanism could be a generic object recognition processor (Bartlett & Searcy, 1993), shared mechanisms can range widely, from low-level visual processing to basic-level object recognition, general cognitive processing speed, and, in the case of temporal measures, even motor response times. This is similar to the concern of others that associations and correlations across object types in tests may merely reflect general performance factors involved in the task, such as cognitive speed, memory, learning or interference effects (Eimer, 2018; Richler, Floyd, & Gauthier, 2015). Indeed, one possible interpretation of the trend to stronger correlations across inverted but not across upright stimuli is that, when more object-specific perceptual mechanisms are reduced in efficiency because of inversion, their relative contribution to performance becomes weaker and that of general performance factors becomes dominant.

Third, for upright stimuli there were fewer significant between-object correlations: in particular, face-search performance did not correlate with performance for any other object type in the upright orientation. This suggests that the perception of upright faces may involve a mechanism distinct from those used by upright words or cars.

It may seem surprising to find correlations between performance for upright words and cars, but not between upright faces and words, or between upright faces and cars. Faces and words are both considered objects of expertise, supported by behavioural observations that both show inversion effects, as well as neuroimaging data showing that both activate similar regions of the fusiform gyrus (Nestor, Plaut, & Behrmann, 2013), a region involved in perceptual expertise (Weiner & Zilles, 2016). They show overlapping patterns of activation that led the *many-to-many hypothesis* to propose that they share and compete for neural resources (Behrmann & Plaut, 2013; Plaut & Behrmann, 2011). This generated neuropsychological predictions that alexic patients would have mild face recognition problems, while prosopagnosic patients would have mild reading difficulties (Behrmann & Plaut, 2013). However, this has not been borne out in studies to date (Robotham & Starrfelt, 2017; Starrfelt, Klargaard, Petersen, & Gerlach, 2016; Susilo, Wright, Tree, & Duchaine, 2015). Our own work and the observations of others has shown that prosopagnosic subjects with right-sided lesions have trouble recognizing handwriting but not with reading (Hills, Pancaroglu, Duchaine, & Barton, 2015; Rentschler, Treutwein, & Landis, 1994) while alexic subjects with left-sided lesions have trouble recognizing facial speech but not face identity (Albonico & Barton, 2017; Campbell et al., 1990; Campbell, Landis, & Regard, 1986). Hence the lack of correlation we found between search performances for face identity versus word identity supports the inference that the processing of word identity and face identity involve different neural resources, which may in part reflect opposing hemispheric specializations. Larger correlational studies that include search for handwriting and facial speech as well as for word and face identity could further explore possible hemispheric differences and relationships.

The fact that the perception of upright faces and cars do not correlate is consistent with longstanding proposals that faces use an orientation-dependent expert mechanism that is not used for other objects, such as cars. There is other evidence that recognition performance is minimally correlated between upright cars and faces in healthy subjects, and that performance on either cars and faces both also dissociate from performance with other object types (McGugin et al., 2012; Richler et al., 2017; Van Gulick et al., 2016). Recent work with factor analysis suggests the existence of a general object cognition factor as well as separate face and vehicle factors, though there is some overlap between all three (Cepulic et al., 2018). However, these studies did not include visual words as an object type.

The correlation between upright words (another type of expertise) and cars (for which we are less expert) is surprising. One might speculate that, as cars and faces are three-dimensional objects with complex surfaces and textures, they may share more with each other than with words, which are two-dimensional line elements. One possible explanation of a relationship between cars and words is that reading is an acquired skill of recent origin that evolved by exploiting general object recognition mechanisms in the human visual system (Dehaene, Cohen, Sigman, & Vinckier, 2005). Perhaps related to this is the idea that the left hemisphere may emphasize part-based processing, which is particularly useful for decoding words and possibly for other object perception (Farah, 1991, 1992; Plaut & Behrmann, 2011). However, further speculation about the reasons for a possible relationship between cars and words should await replication of this finding with other paradigms.

While we did not begin this study with hypotheses about the contrast between target-present and target-absent trials, our preliminary exploration of these effects suggests a few points. First, the presence of false alarms and set-size effects suggests that visual search with our complex objects is consistent with low-threshold theories (Eckstein et al., 2000; Palmer et al., 2000). Second, there may be expertise effects in the mechanisms of search that could be the subject of future studies. While there remains much debate about whether set-size effects are best explained by parallel or serial models of search (Moran, Zehetleitner, Liesefeld, Muller, & Usher, 2016; Townsend & Wenger, 2004; Wolfe, 1998) – and fine-grained analyses can reveal individual differences between subjects for the same stimulus set (Cousineau & Shiffrin, 2004) - our data regarding target-absent/target-present set-size ratios suggest differences in how subjects search the more expert categories of upright words and faces than the less expert category of cars.

There are limitations to this report. While the reliability analyses are promising for the broader goal of creating new tests for assessing the processing of our three different object types, further studies are needed. A larger sample is needed to develop normative data. In particular this is needed to determine the effect of aging. Previous work has demonstrated that aging slows reaction times and reduces the efficiency of visual search (Deary & Der, 2005; Scialfa & Joffe, 1997), and can affect face recognition abilities (Bowles et al., 2009). In our study age did not correlate with the accuracy variables of any test, and, for the response time variables, only correlated with the set-size effect for inverted faces ($r = 0.43$, $p = .037$) and the mean response time for inverted visual words ($r = 0.41$, $p = .045$). This general lack of effect of age may simply reflect the small size of the current sample. However, it is possible is that the object processing reflected in our tests is indeed less affected by aging. There is evidence that face recognition does not decline until after age 70 (Crook & Larrabee, 1992) and that the ability to read remains stable during healthy aging (Gordon, Lowder, & Hoedemaker, 2016).

Other important work for the future would be to gather test-retest data. Stable and replicable scores are desirable characteristics for an experimental test. This is particularly relevant when measuring and estimating individual differences (Hedge, Powell, & Sumner, 2018).

A larger sample would also be important in replicating the results from our correlational analyses. Small samples are associated with increased variability and less reliable correlation coefficients (Schönbrodt & Perugini, 2013). Hence, the interesting pattern of results we obtained must be considered preliminary.

Another issue worth exploring is the effect of stimulus familiarity, which varied among our object types. The words we used are not rare and almost certainly are familiar to all our (literate) subjects, but the faces were anonymous and unfamiliar to all. Familiarity with cars likely fell between these two extremes, with some but not all cars being familiar to some but not all subjects, given variable interest and experience with cars. For faces, there has been recent debate as to whether expertise applies mainly to familiar faces or extends to unfamiliar faces (Rossion, 2018; Young & Burton, 2018). Nevertheless, while performance is better with familiar faces, many effects that are thought to index perceptual expertise for faces, such as the inversion effect and composite face effect, are shown with unfamiliar faces (Rossion, 2018).

In conclusion, object search tests can provide a means of assessing the perceptual processing of different object types, yielding similar accuracy and temporal outcome variables for all. Our results for faces, words and cars suggest that perceptual expertise is reflected in the results, with object set-size effects that are smaller for the more expert categories of words and faces than for cars, inversion effects that are present for words and faces but not cars, and correlation with other measures of perceptual expertise for faces and words, but not for cars. Our preliminary correlational analyses also suggest contributions from object-specific processes that operate on both upright and inverted stimuli, as well as from shared mechanisms that operate with all stimuli in the inverted orientation. They also raise the possibility of a particular contribution from a distinct perceptual mechanism for faces when upright stimuli are processed.

Acknowledgements

This work was supported by Natural Sciences and Engineering Research Council, Discovery Grant RGPIN 319129. JB was supported by Canada Research Chair Tier I, 950-228984, and the Marianne Koerner Chair in Brain Diseases.

References

Aghababian, V., & Nazir, T. A. (2000). Developing normal reading skills: Aspects of the visual processes underlying word recognition. *Journal of Experimental Child Psychology, 76*(2), 123–150. https://doi.org/10.1006/jecp.1999.2540.

Ahlen, E., Hills, C. S., Hanif, H. M., Rubino, C., & Barton, J. J. (2014). Learning to read upside-down: A study of perceptual expertise and its acquisition. *Experimental Brain Research, 232*(3), 1025–1036. https://doi.org/10.1007/s00221-013-3813-9.

Albonico, A., & Barton, J. J. (2017). Face perception in pure alexia: Complementary contributions of the left fusiform gyrus to facial identity and facial speech processing. *Cortex, 96*, 59–72.

Albonico, A., Furubacke, A., Barton, J. J. S., & Oruc, I. (2018). Perceptual efficiency and the inversion effect for faces, words and houses. *Vision Research, 153*, 91–97. https://doi.org/10.1016/j.visres.2018.10.008.

Albonico, A., Malaspina, M., & Daini, R. (2017). Italian normative data and validation of two neuropsychological tests of face recognition: Benton facial recognition test and Cambridge face memory test. *Neurological Sciences, 38*(9), 1637–1643.

Bartlett, J., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25*, 281–316.

Barton, J. J. (2008). Structure and function in acquired prosopagnosia: Lessons from a series of 10 patients with brain damage. *Journal of Neuropsychology, 2*(Pt 1), 197–225.

Barton, J. J. S., Albonico, A., Susilo, T., Duchaine, B., & Corrow, S. L. (2019). Object recognition in acquired and developmental prosopagnosia. *Cognitive Neuropsychology*, 1–31. https://doi.org/10.1080/02643294.2019.1593821.

Barton, J. J., Hanif, H., & Ashraf, S. (2009). Relating visual to verbal semantic knowledge: The evaluation of object recognition in prosopagnosia. *Brain, 132*(12), 3456–3466.

Barton, J. J., Hanif, H. M., Eklinder Björnström, L., & Hills, C. (2014). The word-length effect in reading: A review. *Cognitive Neuropsychology, 31*(5–6), 378–412.

Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends in Cognitive Sciences, 17*(5), 210–219.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., ... Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge face memory test and Cambridge face perception test. *Cognitive Neuropsychology, 26*(5), 423–455. https://doi.org/10.1080/02643290903343149.

Campbell, R., Garwood, J., Franklin, S., Howard, D., Landis, T., & Regard, M. (1990). Neuropsychological studies of auditory-visual fusion illusions. Four case studies and their implications. *Neuropsychologia, 28*(8), 787–802. http://www.ncbi.nlm.nih.gov/pubmed/1701035.

Campbell, R., Landis, T., & Regard, M. (1986). Face recognition and lipreading. A neurological dissociation. *Brain, 109*, 509–521.

Cepulic, D. B., Wilhelm, O., Sommer, W., & Hildebrandt, A. (2018). All categories are equal, but some categories are more equal than others: The psychometric structure of object and face cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(8), 1254–1268. https://doi.org/10.1037/xlm0000511.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain, 125*(5), 1054–1069.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104.

Cousineau, D., & Shiffrin, R. M. (2004). Termination of a visual search with large display size effects. *Spatial Vision, 17*(4–5), 327–352. https://www.ncbi.nlm.nih.gov/pubmed/15559108.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Crook, T. H., 3rd, & Larrabee, G. J. (1992). Changes in facial recognition memory across the adult life span. *The Journals of Gerontology, 47*(3), P138–P141. https://www.ncbi.nlm.nih.gov/pubmed/1573194.

Davies-Thompson, J., Johnston, S., Tashakkor, Y., Pancaroglu, R., & Barton, J. J. (2016). The relationship between visual word and face processing lateralization in the fusiform gyri: A cross-sectional study. *Brain Research, 1644*, 88–97. https://ac.els-cdn.com/S0006899316303602/1-s2.0-S0006899316303602-main.pdf?_tid=029e650d-ddc8-4173-ac9e-462477d6d7a8&acdnat=1536700863_d8973711de78b5cc1473053abdc70712.

Deary, I. J., & Der, G. (2005). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging, Neuropsychology, and Cognition, 12*(2), 187–215.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences, 9*(7), 335–341. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15951224.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107.

Duchaine, B., & Nakayama, K. (2006). The Cambridge face memory test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia, 44*(4), 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001.

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Percept Psychophys, 63*(6), 1004–1013. https://www.ncbi.nlm.nih.gov/pubmed/11578045.

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Percept Psychophys, 62*(3), 425–451. https://www.ncbi.nlm.nih.gov/pubmed/10909235.

Eimer, M. (2018). What do associations and dissociations between face and object recognition abilities tell us about the domain-generality of face processing? *Cognitive Neuropsychology, 35*(1–2), 80–82. https://doi.org/10.1080/02643294.2017.1414691.

Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science, 247*(4943), 721–723. https://www.ncbi.nlm.nih.gov/pubmed/2300824.

Farah, M. J. (1991). Patterns of co-occurrence among the associative agnosias: Implications for visual object recognition. *Cognitive Neuropsychology, 8*, 1–19.

Farah, M. J. (1992). Is an object an object an object? Cognitive and neuropsychological investigations of domain specificity in visual object recognition. *Current Directions in Psychological Science, 1*, 164–169.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble' experts: A framework for studying expert object recognition processes. *Vision research, 38*(15–16), 2401–2428.

Gordon, P. C., Lowder, M. W., & Hoedemaker, R. S. (2016). Reading in normally aging adults. In H. H. Wright (Ed.). *Cognition, language and aging* (pp. 165–191). Philadelphia: John Benjamins Publishing Company.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223–233.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Herzmann, G., Danthiir, V., Schacht, A., Sommer, W., & Wilhelm, O. (2008). Toward a comprehensive test battery for face cognition: Assessment of the tasks. *Behavior Research Methods, 40*(3), 840–857.

Hills, C. S., Pancaroglu, R., Duchaine, B., & Barton, J. J. (2015). Word and text processing in acquired prosopagnosia. *Annals of Neurology, 78*(2), 258–271. https://doi.org/10.1002/ana.24437.

Jenkins, R. (2017). *How many faces do people know? Paper presented at the European conference on visual perception, Berlin, Germany*.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of neuroscience, 17*(11), 4302–4311.

Kleinschmidt, A., & Cohen, L. (2006). The neural bases of prosopagnosia and pure alexia: Recent insights from functional neuroimaging. *Current Opinion in Neurology, 19*(4), 386–391.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud faces database. *Cognition and Emotion, 24*(8), 1377–1388.

Leff, A., Spitsyna, G., Plant, G., & Wise, R. (2006). Structural anatomy of pure and hemianopic alexia. *Journal of Neurology, Neurosurgery & Psychiatry, 77*(9), 1004–1007.

Macmillan, N., & Creelman, C. (1991). *Detection theory: A user's guide.* New York: Cambridge University Press.

McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012). The Vanderbilt Expertise Test reveals domain-general and domain-specific sex effects in object recognition. *Vision Research, 69*, 10–22. https://doi.org/10.1016/j.visres.2012.07.014.

Moran, R., Zehetleitner, M., Liesefeld, H. R., Muller, H. J., & Usher, M. (2016). Serial vs. parallel models of attention in visual search: Accounting for benchmark RT-distributions. *Psychonomic Bulletin & Review, 23*(5), 1300–1315. https://doi.org/10.3758/s13423-015-0978-1.

Nestor, A., Plaut, D. C., & Behrmann, M. (2013). Face-space architectures: Evidence for the use of independent color-based features. *Psychological Science, 24*(7), 1294–1300.

Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception, 22*(11), 1287–1298. https://doi.org/10.1068/p221287.

Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for measurement of change. *Psychological Bulletin, 82*(1), 85–86.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40*(10–12), 1227–1268. https://www.ncbi.nlm.nih.gov/pubmed/10788638.

Pilz, K. S., Thornton, I. M., & Bulthoff, H. H. (2006). A search advantage for faces learned in motion. *Experimental Brain Research, 171*(4), 436–447. https://doi.org/10.1007/s00221-005-0283-8.

Plaut, D. C., & Behrmann, M. (2011). Complementary neural representations for faces and words: A computational exploration. *Cognitive Neuropsychology, 28*(3–4), 251–275. https://doi.org/10.1080/02643294.2011.609812.

Rentschler, I., Treutwein, B., & Landis, T. (1994). Dissociation of local and global processing in visual agnosia. *Vision Research, 34*, 963–971.

Richler, J. J., Floyd, R. J., & Gauthier, I. (2015). About-face on face recognition ability and holistic processing. *Journal of Vision, 15*(9), 15. https://doi.org/10.1167/15.9.15.

Richler, J. J., Wilmer, J. B., & Gauthier, I. (2017). General object recognition is specific: Evidence from novel and familiar objects. *Cognition, 166*, 42–55. https://doi.org/10.1016/j.cognition.2017.05.019.

Robotham, R. J., & Starrfelt, R. (2017). Face and word recognition can be selectively

affected by brain injury or developmental disorders. *Frontiers in Psychology, 8*, 1547. https://doi.org/10.3389/fpsyg.2017.01547.
Rossion, B. (2018). Humans are visual experts at unfamiliar face recognition. *Trends in Cognitive Sciences, 22*(6), 471–472. https://doi.org/10.1016/j.tics.2018.03.002.
Russell, R., Chatterjee, G., & Nakayama, K. (2012). Developmental prosopagnosia and super-recognition: No special role for surface reflectance processing. *Neuropsychologia, 50*(2), 334–340. https://doi.org/10.1016/j.neuropsychologia.2011.12.004.
Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612.
Scialfa, C. T., & Joffe, K. M. (1997). Age differences in feature and conjunction search: Implications for theories of visual search and generalized slowing. *Aging, Neuropsychology, and Cognition, 4*(3), 227–246.
Sekuler, A., Gaspar, C., Gold, J., & Bennett, P. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology, 14*, 391–396.
Sheldon, C. A., Abegg, M., Sekunova, A., & Barton, J. J. (2012). The word-length effect in acquired alexia, and real and virtual hemianopia. *Neuropsychologia, 50*(5), 841–851. https://doi.org/10.1016/j.neuropsychologia.2012.01.020.
Starrfelt, R., Klargaard, S. K., Petersen, A., & Gerlach, C. (2016). Are reading and face processing related? An investigation of reading in developmental prosopagnosia. bioRxiv, 039065.
Susilo, T., Wright, V., Tree, J. J., & Duchaine, B. (2015). Acquired prosopagnosia without word recognition deficits. *Cognitive Neuropsychology, 32*(6), 321–339. https://doi.org/10.1080/02643294.2015.1081882.
Tong, F., & Nakayama, K. (1999). Robust representations for faces: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 25*(4), 1016–1035. https://www.ncbi.nlm.nih.gov/pubmed/10464943.
Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review, 11*(3), 391–418. https://www.ncbi.nlm.nih.gov/pubmed/15376788.
Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136. https://www.ncbi.nlm.nih.gov/pubmed/7351125.
Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471–491.
Van Gulick, A. E., McGugin, R. W., & Gauthier, I. (2016). Measuring nonvisual knowledge about object categories: The semantic Vanderbilt expertise test. *Behav Res Methods, 48*(3), 1178–1196. https://doi.org/10.3758/s13428-015-0637-5.
Warrington, E. K. (1984). *Recognition memory test.* UK: Windsor.
Weiner, K. S., & Zilles, K. (2016). The anatomical and functional specialization of the fusiform gyrus. *Neuropsychologia, 83*, 48–62.
Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research, 34*(9), 1187–1195. https://www.ncbi.nlm.nih.gov/pubmed/8184562.
Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*(1), 33–39.
Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences, 7*(2), 70–76. https://www.ncbi.nlm.nih.gov/pubmed/12584025.
Wolfe, J. M. (2019). Visual attention: The multiple ways in which history shapes selection. *Current Biology, 29*(5), R155–R156. https://doi.org/10.1016/j.cub.2019.01.032.
Yang, N., Shafai, F., & Oruc, I. (2014). Size determines whether specialized expert processes are engaged for recognition of faces. *Journal of Vision, 14*(8), 17. https://doi.org/10.1167/14.8.17.
Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141–145.
Young, A. W., & Burton, A. M. (2018). Are we face experts? *Trends in Cognitive Sciences, 22*(2), 100–110. https://doi.org/10.1016/j.tics.2017.11.007.
Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language, 93*(3), 369–373. https://doi.org/10.1016/j.bandl.2004.10.010.