Experience moderates overlap between object and face recognition, suggesting a common ability

Isabel Gauthier Department of Psychology, Vanderbilt University, Nashville, TN, USA

Rankin W. McGugin Department of Psychology, Vanderbilt University, Nashville, TN, USA

Jennifer J. Richler Department of Psychology, Vanderbilt University, Nashville, TN, USA

Grit Herzmann Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA

Magen Speegle Department of Psychology, Vanderbilt University, Nashville, TN, USA

Ana E. Van Gulick Department of Psychology, Vanderbilt University, Nashville, TN, USA

Some research finds that face recognition is largely independent from the recognition of other objects; a specialized and innate ability to recognize faces could therefore have little or nothing to do with our ability to recognize objects. We propose a new framework in which recognition performance for any category is the product of domain-general ability and category-specific experience. In Experiment 1, we show that the overlap between face and object recognition depends on experience with objects. In 256 subjects we measured face recognition, object recognition for eight categories, and self-reported experience with these categories. Experience predicted neither face recognition nor object recognition but moderated their relationship: Face recognition performance is increasingly similar to object recognition performance with increasing object experience. If a subject has a lot of experience with objects and is found to perform poorly, they also prove to have a low ability with faces. In a follow-up survey, we explored the dimensions of experience with objects that may have contributed to self-reported experience in Experiment 1. Different dimensions of experience appear to be more salient for different categories, with general self-reports of expertise reflecting judgments of verbal knowledge about a category more than judgments of visual performance. The complexity of experience and current limitations in its measurement support the importance of aggregating across multiple categories. Our findings imply that both face and object recognition are supported by a common, domain-general ability expressed through experience with a category and best measured when accounting for experience.

Introduction

It is widely believed that face recognition constitutes a special ability that is relatively independent from domain-general factors such as general cognitive ability, motivation, or attention (Wilhelm et al., 2010; Wilmer et al., 2010; Zhu et al., 2010). Additionally, face recognition ability does not tend to correlate very strongly with the recognition of nonface objects (Dennett, McKone, Edwards, & Susilo, 2012; Wilhelm et al., 2010). Here, we present evidence challenging this distinction. Our results imply that face and object recognition share much more variance than originally thought and reflect an underlying ability that generally supports the acquisition of skills in discriminating visually similar objects. Evidence for such a domain-

general ability would have broad-ranging implications. Beyond influencing aptitude in social interactions, this ability could constrain performance in domains such as learning to play chess, identifying tumors in X-rays or magnetic resonance imaging pictures, identifying fingerprints, or reading airport security displays. Expertise in some of these domains is associated with processing strategies and neural responses similar to those found for faces (Gauthier et al., 2000, 2002; Harley et al., 2009; McGugin, Gatenby, Gore, & Gauthier, 2012; McGugin, Van Gulick, Tamber-Rosenau, Ross, & Gauthier, in press), but such results are not sufficient to test hypotheses regarding the correlation between face and object recognition abilities across subjects. Because many domains of human activity surely benefit from skilled perception, measuring individual differences that may explain variability in high-level visual cognition (beyond what can be accounted for by general cognitive ability) could greatly increase predictions about performance.

Although many agree that face recognition constitutes an important skill, the study of individual differences in face recognition in the normal adult population with consideration to the psychometric properties of the measures is surprisingly recent. This growing literature has established the Cambridge Face Memory Test (CFMT) as a measure of face recognition ability (Duchaine & Nakayama, 2006). The CFMT successfully discriminates individuals over a wide range of performance (Bowles et al., 2009; Germine, Duchaine, & Nakayama, 2011; Russell, Duchaine, & Nakayama, 2009). Consistent with other measures of face recognition (Wilhelm et al., 2010), CFMT scores are generally independent from (or at least not linearly correlated with) verbal memory (Bowles et al., 2009; Wilmer et al., 2010) and intelligence quotient (Davis et al., 2011).

In a study that found face recognition to be a highly heritable ability, Wilmer et al. (2010) also assessed the independence of face recognition from object recognition. In a very large sample ($n = 3,004$), the comparison of the CFMT with a similar test of recognition for abstract art yielded a low correlation ($r_{3002} = 0.26$), accounting for less than 7% ($R^2 = 0.07$) of the variance in the CFMT. Likewise, Dennett et al. (2011) designed the Cambridge Car Memory Test, which accounted for only 13.6% of the variance in the CFMT ($n = 142$). As a result, the authors concluded that face and object processing are largely independent. Others have compared performance for faces and a single category of nonface objects with similar results (Dennett et al., 2012; Wilhelm et al., 2010).

Before accepting the independence of face and object recognition abilities based on limited shared variance between performance for faces and objects in a single nonface domain, however, an alternative has to be considered. Face and object recognition may be expressions of a single domain-general ability that allows people to discriminate visually similar objects, and this domain-general ability is best expressed for individuals in domains with which they have sufficient experience. Such a domain-general ability is not inconsistent with the above-mentioned results. First, shared variance around 14% reflects correlations in the 0.3 to 0.4 range, which is comparable with the magnitude of correlations obtained between individual tasks that are not considered to reflect independent abilities. For instance, different tasks measuring working memory or short-term memory skills can show correlations of this magnitude, even though each of the task's loading on latent factors that represent these constructs can be much higher ($\sim$0.6–0.8; Engle, Tuholski, Laughlin, & Conway, 1999). Second, even if there is a domain-general ability, performance with faces and nonface objects may not be strongly correlated since there is more variability in experience with objects relative to faces. On the one hand, in a domain such as face recognition, where most people are highly motivated to excel and have extensive opportunity to practice, performance likely expresses this ability in full. In domains such as cars or abstract art, on the other hand, motivation and experience vary greatly, so performance will not reflect innate ability alone but rather the product of category-specific experience and domain-general ability. In fact, it is plausible that people vary more in their experience with objects than they do in their general visual ability, in which case performance with objects would mostly reflect experience.

Here, we reassess the shared variance between face and object recognition and provide the first demonstration that expressing this ability depends on experience. Experts with extensive real-world experience individuating objects in nonface categories (e.g., birds, cars) also demonstrate a variety of behavioral (Curby, Glazek, & Gauthier, 2009; McGugin, McKeeff, Tong, & Gauthier, 2010) and neural (Gauthier, Curran, Curby, & Collins, 2003; Harley et al., 2009; Xu, 2005) markers of face perception in their domain of expertise. We hypothesize that subjects with more individuating experience will show more similar performance for faces and nonface objects than those with less individuating experience. Using many object categories is important in this framework because we assume that performance with a category with which a subject has modest experience provides very little information about their ability. To the extent that experience between different categories is moderately correlated (e.g., this correlation across all subjects is $r = 0.38$ in our Experiment 1), using several categories has at least two advantages: 1) the standard benefits of aggregating to reduce noise (Rushton, Brainerd, & Pressley, 1983), and 2) any subject is more likely to

have experience with at least some of these categories, providing us with more information about their ability.

Virtually no work has been done in the measurement of experience with objects (see Stanovich & Cunningham, 1992, for research on measuring exposure to print as a contributor to verbal skills). Since the self-ratings used in Experiment 1 proved to be a significant moderator of object performance, Part 2 offers an analysis of a survey designed to explore what dimensions subjects may have emphasized when answering these self-rating questions.

Experiment 1

Methods

Subjects

A total of 256 individuals (130 males, mean age = 23.7 years, $SD = 4.34$; 126 females, mean age = 22.6 years, $SD = 4.03$) participated for a small honorarium or course credit. Subjects were recruited as part of other experiments (samples from McGugin, Gatenby et al., 2012; McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012; Van Gulick, McGugin, & Gauthier, 2012), none of which included the analyses reported here.[1] All subjects had normal or corrected-to-normal visual acuity and provided written informed consent. The experiment was approved by the Institutional Review Board at Vanderbilt University.

Cambridge Face Memory Test

In the CFMT (Duchaine & Nakayama, 2006), subjects studied six target faces and completed an 18-trial introductory learning phase, after which they were presented with 30 forced-choice test displays. Subjects had to select the target among two distractors on each trial. The matching faces varied from their original presentation in terms of lighting condition, pose, or both. Next, subjects were again presented with the six target faces to study, followed by 24 trials presented in Gaussian noise.

Vanderbilt Expertise Test

The Vanderbilt Expertise Test (VET; McGugin, Richler et al., 2012) is similar to the CFMT in format and in psychometric properties. It includes 48 items for each of eight categories of objects (butterflies, cars, leaves, motorcycles, mushrooms, owls, planes, wading birds; see Supplementary Material and Supplementary Table S1).

Before the start of a category-specific block, subjects viewed a study screen with one exemplar from each of six species or models. For the first twelve trials, one of the studied exemplars was presented with two distractors from another species or model in a forced-choice paradigm with feedback (Figure 1). Then, subjects reviewed the study screen and were warned that in the subsequent 36 trials there would be no feedback and that target images would be different exemplars of the studied species or models, requiring generalization across viewpoint, size, and backgrounds. Subjects viewed image triplets and indicated which exemplar corresponded to one of the target species or models studied. For a complete description of the VET, see McGugin, Richler et al. (2012).

Self-rating of experience

Subjects rated themselves on their expertise with each of the eight VET categories and with faces on a 9-point scale, considering "interest in, years of exposure to, knowledge of, and familiarity with each category." To be clear, here and in past work, even though the question is phrased in terms of "expertise," it should not be assumed that these ratings correspond to *perceptual* skills, especially because such ratings tend not to be strongly related to perceptual performance (Barton, Hanif, & Ashraf, 2009; McGugin, Richler et al., 2012). For clarity, we therefore refer to this measure as one of "experience" (a conjecture supported by Part 2). We were able to retest a small number of subjects ($n = 39$) an average of 719 days ($SD = 203$) after the original test. Test–reretest reliability (Pearson r) for each category was as follows: faces, 0.20, 95% CI [−0.12, 0.48]; butterflies, 0.56, 95% CI [0.30, 0.74]; cars, 0.59, 95% CI [0.34, 0.76]; leaves, 0.66, 95% CI [0.44, 0.81]; motorcycles, 0.63, 95% CI [0.39, 0.79]; mushrooms, 0.63, 95% CI [0.39, 0.79]; owls, 0.54, 95% CI [0.27, 0.73]; planes, 0.63, 95% CI [0.39, 0.79]; and wading birds, 0.39, 95% CI [0.09, 0.63]. The test–retest reliability of the ratings across nonface categories was 0.69, 95% CI [0.48, 0.83], and with faces included it was 0.80, 95% CI [0.65, 0.89], because even though the reliability of the face ratings is low, the ratings are consistently higher than ratings for the other categories. While the test–retest reliability of these self-reports was low for several of these categories, it reached 0.60, 95% CI [0.35, 0.77] for the measure we use in our main analyses, O-EXP (the summed ratings for all nonface categories). As a comparison, one large sample study reports a test–retest reliability for the CFMT of 0.70 (Wilmer et al., 2010).

Results and discussion

Our analyses consider the relationship between face recognition (CFMT), object recognition (O-PERF; an
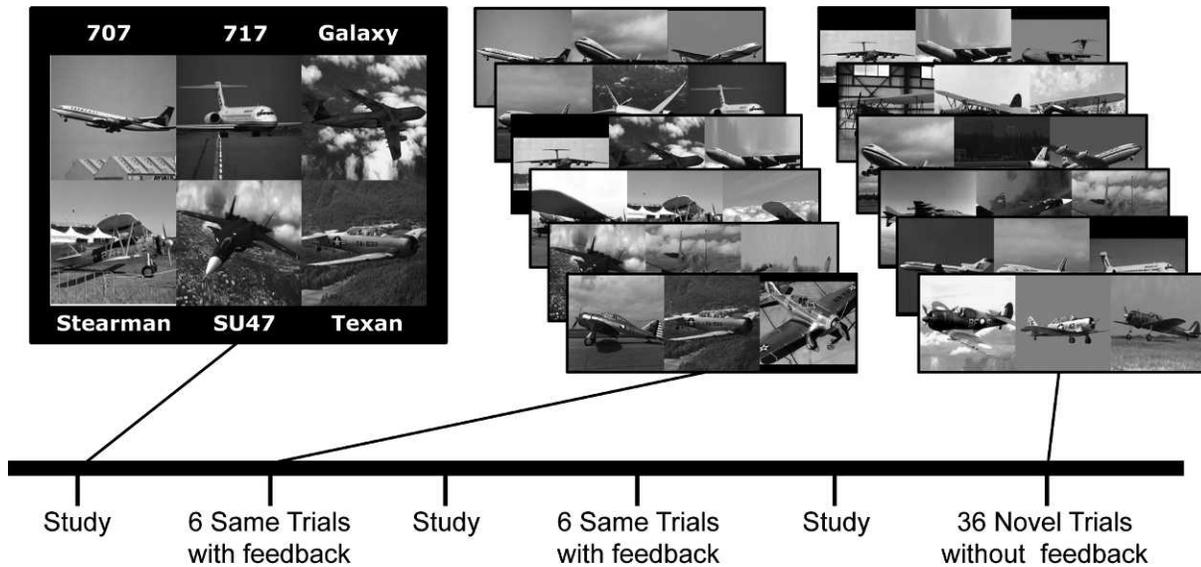
Figure 1. Task structure for a single category of the VET.

aggregate score of performance for all eight object categories), and experience with these eight categories (O-EXP; an aggregate score of the self-ratings for all eight categories). The mean shared variance in object performance among any two categories was 12% (ranging from 0% for cars and leaves to 31% for leaves and butterflies). The mean shared variance in self-ratings among categories was 14.7% (ranging from 10.1% to 18.4%).

As observed previously (Barton et al., 2009; McGugin, Gatenby et al., 2012; McGugin, Richler et al., 2012), self-ratings of experience for each category were generally not highly predictive of performance (the largest amount of shared variance—R^2 expressed as a percentage—between performance and self-ratings was 14.6% for cars; see Supplementary Table S1). Here, we used self-ratings to estimate what self-reports most likely tap into: not recognition *performance* as much as *experience with* the object categories. A recent review of several meta-analyses suggested that correlation between people's predictions of abilities across a wide range of measures (e.g., intellectual, sports, memory) is generally limited ($r = 0.29$ or 8.4% shared variance; Zell & Krizan, 2014). It is possible that insight into performance is particularly low in perceptual domains because people rarely have the opportunity to compare their perceptual skills with those of others. However, we surmise that they may have better knowledge of their *experience* with a category relative to that of other people.

Remember that face recognition has been deemed to be independent of object recognition because its shared variance with car recognition was no more than 14% (e.g., Dennett et al., 2012). However, this claim did not consider more than a single object category. Again, the average correlation between any two nonface object categories is $r = 0.34$, which translates to a shared variance of 12%, 95% CI [1%, 20%] (see Table 1). This is the case, even though the reliability (estimated by Cronbach alpha) is above 0.7 for the majority of these categories. By this approach, any of these object categories could have been deemed to be independent from "object recognition." This highlights the importance of using more than a single category not only to measure object recognition ability but also to interpret it.

Theoretically, there is a latent factor of general visual ability underlying all category-specific measures of performance, but in practice its measurement is complicated by variability in experience that contributes to performance for each category. The shared variance in experience across different categories (14.7%) suggests that there could be a domain-general aspect of experience, whereby some people are more disposed than others to learn to individuate objects in various domains. We come back to this in Part 2.

O-PERF and CFMT showed a significant correlation ($r_{254} = 0.357$, $p = 0.0026$, shared variance 12.7%) of magnitude similar to that found when only cars were used in prior work (Dennett et al., 2012). O-EXP was correlated with neither the CFMT ($r_{254} = -0.021$, $p = 0.74$) nor O-PERF ($r_{254} = -0.025$, $p = 0.69$).

We then performed a multiple regression on CFMT scores, entering O-PERF and O-EXP and their interaction simultaneously as predictors. Externally studentized residuals were examined. One outlier (Cook's $D = 0.12$) was removed, although none of the results are qualitatively affected. As can be seen in Table 2, not only was O-PERF a significant predictor of CFMT, but, critically, its interaction with O-EXP was a significant predictor of CFMT. Subject sex does not modulate the experience–performance interaction

Task	1	2	3	4	5	6	7	8	9	Cronbach α
1. CFMT	–									0.90
2. VET butterflies	0.32	–								0.77
3. VET cars	0.24	0.11	–							0.84
4. VET leaves	0.27	0.56	0.02	–						0.68
5. VET motorcycles	0.19	0.39	0.25	0.35	–					0.73
6. VET mushrooms	0.24	0.43	0.09	0.40	0.33	–				0.62
7. VET owls	0.26	0.49	0.07	0.47	0.36	0.40	–			0.75
8. VET planes	0.15	0.41	0.33	0.39	0.45	0.32	0.33	–		0.81
9. VET wading birds	0.22	0.43	0.09	0.49	0.38	0.38	0.49	0.43	–	0.68
O-PERF (VET all categories)										0.92

Table 1. Correlations among performance measures, and their reliability.

when accounting for variability in the CFMT ($p = 0.16$; see Supplementary Table S2).

To unpack this interaction, Figure 2A displays the correlation between CFMT and O-PERF separately for subjects in three terciles of O-EXP. Shared variance is near zero in subjects with very little experience with our object categories, whereas a robust correlation is obtained in subjects with more experience. Figure 2B shows the same effect for the entire sample, in smaller bins of O-EXP, to illustrate how the effect grows monotonically with experience. Because the first sample (O-EXP ≤ -1.5 SD) has only 10 subjects, we include a sample of a similar size ($n = 9$) at the other extreme of O-EXP (>2 SD), illustrating a very dramatic difference even with such small samples. For subjects with O-EXP more than 2 SD above the mean (seven men, two women), the shared variance between CFMT and O-PERF is 59% ($p = 0.01$), with a value corrected for reliability (correction for attenuation; Spearman, 1904) based on our Cronbach alpha for CFMT and VET of 73%, 95% CI [49.5, 100]. Even though this small sample estimate has a large confidence interval, it illustrates how with increasing experience the results are less consistent with the suggestion that face recognition is independent from object recognition. According to our findings, experience with objects appears to be necessary for performance with objects to reflect the same ability measured by the CFMT. If a subject has a lot of experience with objects and is found to perform poorly, then we can expect that they have a low ability (including with faces). But if they perform poorly and have little experience, the performance score is not a good measure of ability and may reflect other sources

of variance supporting strategies that subjects may use when they have little knowledge of a category's diagnostic dimensions.

Interestingly, Cho et al. (submitted) recently provided the first test of the assumption that CFMT measures a single source of variance (unidimensionality assumption) in a large sample ($n = 2,497$). While there was some evidence for multidimensionality, 71% of the variance explained by CFMT scores was due to a single factor onto which all trials loaded and the rest was divided among three roughly equal factors that appeared to relate to different CFMT target faces. In other words, there was little evidence for any variance that was not related to face recognition, and certainly not of sufficient magnitude to explain the present relationships with object performance at high levels of experience.

Dennett et al. (2012) performed an analysis similar to ours, although with a single object category and using a measure of knowledge with cars instead of a self-report of experience. They found no interaction with car recognition when predicting the CFMT. This result is not inconsistent with ours: We also do not find a significant experience–performance interaction when predicting the CFMT with the VET for most individual categories, the exception being motorcycles (Supplementary Table S2). Thus, our findings may depend on the well-known value of aggregation (Rushton et al., 1983), especially for the experience measure. There are several possibilities that could explain why we observe an interaction between experience and performance in predicting CFMT scores when we use eight categories but not each individual category. There may be several distinct dimensions that influence a subject's self-ratings of experience on a given

Model and predictor	Beta	SE	t	p
CFMT (R^2 adjusted = 0.15)				
Intercept	0.801855	0.007	116	<0.0001
z O-EXP	−0.000866	0.007	−0.124	<0.901
z O-PERF	0.044406	0.007	6.32	<0.0001
z O-PERF × z O-EXP	0.016399	0.006	2.65	0.009

Table 2. Results of multiple regression on CFMT scores with O-PERF, O-EXP, and O-PERF × O-EXP entered simultaneously as predictors.
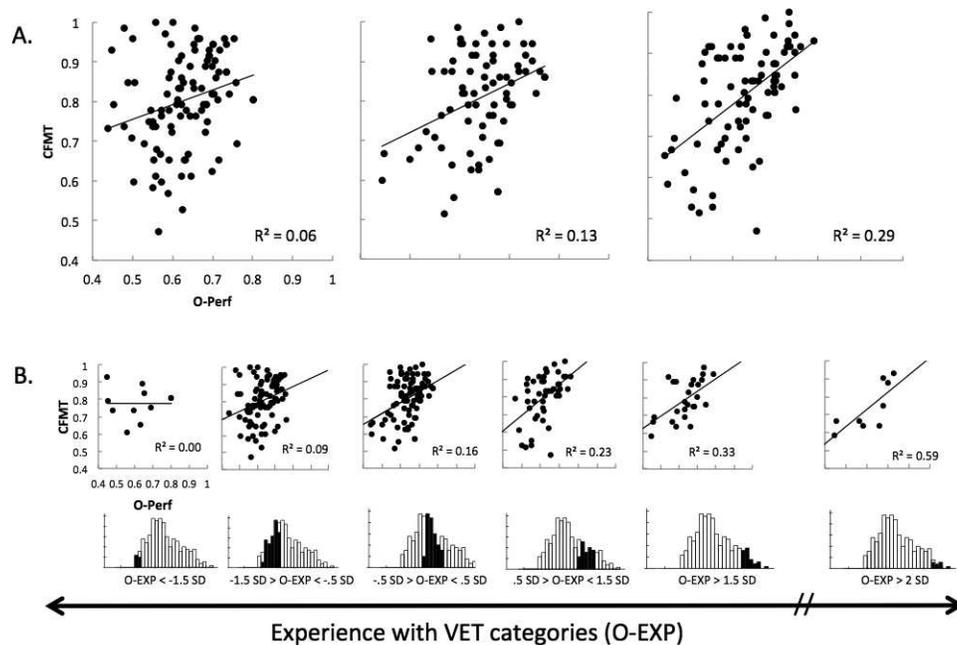
Figure 2. Correlation between CFMT and O-PERF as a function of O-EXP. (A) Correlation for subjects in three terciles of O-EXP. (B) The correlation plotted for subgroups of the sample with O-EXP scores less than −1.5 SD, between −1.5 and −0.5 SD, between −0.5 and 0.5 SD, between 0.5 and 1.5 SD, more than 1.5 SD, and more than 2 SD.

category, and if only one of these dimensions is relevant as a moderator of performance, the self-rating variance for each category may be dominated by a combination of irrelevant factors rather than by the correct dimension. Averaging across categories would pull out this relevant dimension only to the extent that the irrelevant dimensions are correlated across categories. A variation on this account is that the effect is small for each category, with subjects varying on irrelevant dimensions that are not correlated across categories, for *both* visual performance and experience such that the correlation emerges only with aggregation. Simulations of the present effects in a neural network of object recognition, with ability mapped onto the number of hidden units available for learning and experience mapped onto the number of training epochs with nonface objects, suggests that effects are not seen for individual categories with 256 simulated subjects but are found in individual categories with more subjects (800 in this case; Wang, Gauthier, & Cottrell, submitted). Yet another possibility is that aggregating self-report across categories isolates a domain-general aspect of experience (i.e., some people are more interested in individuating objects, regardless of domain), and this is the dimension that mediates the correlation between face and object recognition performance.

In Part 2, we conducted a survey investigating experience with objects to explore the nature of the self-ratings of experience used in Experiment 1. This exercise is exploratory, but with Experiment 1 demonstrating the important role of experience in interpreting any

relationship between object and face recognition performance, we have to acknowledge that the measure of experience used in Experiment 1 was very crude. For all the convenience of our measure being a single question, it has not been validated against any other measures. There is no published study focusing on the measurement of experience with objects, and while we do not expect to settle this question here, we wanted to provide a first effort at clarifying what was measured by the series of category-specific questions in Experiment 1.

Part 2: Experience survey

We aimed to shed some light on how subjects may have interpreted the general question used for self-ratings with object categories in Experiment 1. Our aim was not to create a standard instrument for self-ratings of experience or to test a specific theory of experience with objects, so we limit our analyses to those most relevant to the interpretation of Experiment 1.

Methods

Subjects

Subjects were recruited in Amazon Mechanical Turk (Amazon Web Services, Amazon Inc., Seattle, WA). A total of 197 subjects (81 males, mean age = 33.2 years, $SD = 13.5$; 116 females, mean age = 35.9 years, $SD =$

14.2) completed the experiment (an additional five subjects were dropped because they failed to answer some of the questions). Data were collected and managed using REDCap electronic data capture tools (Harris et al., 2009) hosted at Vanderbilt University.

Experience questionnaire

Domain-general questions: The first four questions sought information about subjects' general experience with individuating objects, before they were asked about specific categories. These questions were as follows: (1) Generally speaking, how strong is your interest in classifying objects in their various subcategories (e.g., learning about different kinds of insects, plants, shoes, tools, etc.)?; (2) Generally speaking, how easily do you learn to recognize objects visually?; (3) Generally speaking, relative to the average person, how much of your time at work or school involves recognizing things visually?; and (4) Generally speaking, relative to the average person, how much of your free time involves recognizing things visually? In each case a rating scale from 1 to 9 was used, with appropriate labels shown for 1 and 9 (e.g., not strong at all/very strong; much less than average/much more than average).

Omnibus category-specific self-ratings: Next, we asked the self-rating questions used in Experiment 1 ("Please rate yourself on your expertise with X, considering your interest in, years of exposure to, knowledge of, and familiarity with this category") in the same order as in Experiment 1 (butterflies, cars, leaves, motorcycles, mushrooms, owls, planes, wading birds), using a scale from 1 to 9.

Category-specific experience questions: Finally, we asked the following six category-specific questions (each question was presented for each category in alphabetical order before moving to the next question): (1) If you were asked to write an essay about different kinds of X, how extensive and detailed do you think your essay would be? (1: very short and simple; 9: very long and detailed); (2) If you saw a specific X in a TV show, how sure are you that you could recognize that item among similar images if you were tested the next day? (1: not sure at all; 9: absolutely sure); (3) How important is the domain of X to you, relative to all the other things you are interested in? (1: one of the least important; 9: one of the most important); (4) How often do you read text (books, magazines, or online) that contains information about X? (1: almost never; 9: almost every week); (5) How often do you look at images of X in movies, television, or other kinds of documents (books, magazines, or online)? (1: almost never; 9: almost every week); and (6) If you are interested in objects of the category X, when did this interest begin? (I have no real interest in this category; 1

year ago or less; 2 to 3 years ago; 4 to 5 years ago; 6 or more years ago). These questions were presented in two different orders for two different groups of subjects in order to mitigate order effects (96 subjects saw the questions in the order listed here, and 101 subjects saw the questions in the reverse order).

Importantly, the domain-general questions were answered before any of the category-specific questions so that if domain-general information could account for answers in category-specific ratings it would not be because these categories had been made salient. Likewise, the self-rating questions used in Experiment 1 was answered for all categories before we asked any of the more detailed domain-specific questions so that if these more detailed questions account for the domain-specific ratings it would not be because these dimensions had been made salient.

Results

Even though mean age was higher in the survey sample than in Experiment 1, the pattern of response for the category-specific ratings of expertise was similar. The average correlation between pairwise ratings was 0.45 in the survey, similar to 0.38 in Experiment 1, and the correlation between the 28 pairwise correlations across the two studies was 0.78 ($p < 0.0001$). Because our analyses are exploratory, we are more interested in patterns of results than in statistical significance and mainly focus on effect sizes (*r*-values). It is important to keep in mind that the test–retest reliabilities for the omnibus questions for several of these categories, as measured in Experiment 1, were low. As such we believe that the omnibus question will not suffice as a reliable measure of experience in future work.

Accounting for O-EXP using domain-general questions

To what extent can the domain-general questions account for the aggregate O-EXP score used in Experiment 1? Table 3 presents a regression on O-EXP using all four predictors entered simultaneously as well as the zero-order correlations (those correlations among only the two variables). Both analyses suggest that the domain-general component in O-EXP may be best accounted for by subjects' interest in identifying objects (and not their estimate of their ability to do so) and the relevance of this type of activity to their work.

However, we can account for almost as much variance in O-EXP using a single predictor that is the average of the four domain-general questions (R^2 adjusted = 0.14 vs. 0.15 for four predictors). For subsequent analyses we use the average of the four general questions.

Model and predictor	Beta	SE	t	p	Zero-order r
O-EXP (R^2 adjusted = 0.15)					
Intercept	0.757461	0.4791	1.58	0.1155	
Q1. Interest	0.167538	0.0503	3.33	0.001	0.32
Q2. Visual ability	−0.010923	0.0678	−0.161	0.8722	0.18
Q3. Work experience	0.183639	0.0546	3.36	0.0009	0.32
Q4. Hobby experience	0.045136	0.0604	0.747	0.4558	0.23

Table 3. Zero-order correlations between answers on the four domain-general questions and O-EXP, and multiple regression on O-EXP using these four predictors.

Accounting for O-EXP using category-specific questions

To what extent do the category-specific questions account for the aggregate O-EXP score used in Experiment 1? When we ask specific questions about experience rather than one omnibus question, the answers are moderately correlated across categories (between 0.2 and 0.5; Table 4). Moreover, the average for each of these questions across categories correlates moderately with O-EXP (between 0.4 and 0.7), which is numerically higher in all cases than the correlation with the average of the domain-general questions (between 0.1 and 0.3; see Table 4).

A multiple regression on O-EXP using six simultaneously entered predictors (the average across categories for each question) accounted for 55% of the variance, with all the predictors significant except for visual memory. A stepwise regression adding the average of the domain-general questions to this model accounted for only an additional 4% of the variance (although it was significant).

Thus, to the extent that O-EXP can be explained by domain-general factors, those factors may be better expressed by an aggregate of different dimensions of experience for the categories we happened to use than by domain-general ratings. Caveats include the fact that we may have better sampled dimensions of experience with our six domain-specific questions relative to our four domain-general questions and that each index of a dimension of experience across categories already benefits from aggregating over eight different measurements.

	Average r	SD	r with O-EXP	r with GEN
Q1. Essay	0.51	0.15	0.54	0.22
Q2. Visual memory	0.47	0.16	0.39	0.15
Q3. Importance	0.36	0.18	0.64	0.30
Q4. Text	0.34	0.18	0.68	0.30
Q5. Images	0.32	0.17	0.59	0.24
Q6. Duration	0.24	0.20	0.47	0.19

Table 4. Average correlation between the ratings on each question across the eight object categories; correlation of average ratings across categories for each question and O-EXP and GEN (the average of the four domain-general questions).

Accounting for category-specific self-ratings of experience

To what extent do the six specific dimensions and the average of the four domain-general questions account for variance in the category-specific rating questions used in Experiment 1 for each category?

The zero-order correlations for the questions are presented in Table 5 (in descending order of their average correlation) and suggest that each of these factors shares nonnegligible variance with the self-rating questions used in Experiment 1, including the average of the domain-general ratings, although it produced the lowest correlations. These results provide validity to the omnibus self-rating questions: They do appear to tap into a number of dimensions of experience with objects.

Table 6 shows the results of multiple regressions (one for each category, with questions shown in descending order of their average correlations) in which all category-specific predictors and the general average were entered simultaneously to predict each of the category-specific self-ratings.

Interestingly, some of the predictors provide limited unique variance accounting for the self-ratings. In particular, the question about visual memory was not a significant predictor for any of the categories, consistent with self-ratings in Experiment 1 not being strong predictors of VET performance. In contrast, self-reports of verbal knowledge accounted for unique variance for most categories (seven out of eight) and is a candidate for the dimension that was expressed by aggregation in Experiment 1. It would be interesting to measure verbal knowledge in specific domains, rather than self-report, to estimate experience. Also interesting is the finding that the average of the domain-general questions produced a significant partial correlation for five of the categories. Domain-general information is another candidate for a common factor that is expressed by aggregation.

The results suggest that even though the omnibus self-rating question in Experiment 1 was phrased the same way for all categories, people may attend to different aspects of their experience depending on the category. For instance, frequency of looking at images (question five) provided the most unique variance for

	BUT	CAR	LEA	MOT	MUS	OWL	PLA	WAD	Average
Q3. Importance	0.53	0.67	0.51	0.72	0.60	0.55	0.62	0.61	0.60
Q4. Text	0.53	0.69	0.59	0.66	0.57	0.58	0.60	0.55	0.60
Q1. Essay	0.49	0.66	0.53	0.63	0.51	0.48	0.60	0.54	0.55
Q5. Images	0.49	0.53	0.59	0.57	0.55	0.58	0.51	0.63	0.55
Q6. Duration	0.45	0.56	0.43	0.50	0.41	0.47	0.60	0.46	0.48
Q2. Visual memory	0.23	0.56	0.42	0.55	0.45	0.42	0.53	0.49	0.45
GEN average	0.44	0.19	0.29	0.26	0.21	0.26	0.25	0.31	0.28

Table 5. Zero-order correlations of the ratings on each of the detailed category-specific questions and the average of the four domain-general questions with the self-rating question for each category. The critical two-tail *r*-value is 0.14. *Notes*: GEN average, average of general questions; BUT, butterflies; CAR, cars; LEA, leaves; MOT, motorcycles; MUS, mushrooms; OWL, owls; WAD, wading birds.

leaves but none for planes, while duration of interest (question six) provided the most unique variance for planes but none for motorcycles. We would be cautious about interpreting any of these specific effects without replication, and it should be kept in mind that there may be other dimensions of experience that we did not sample here. But this variability suggests that aggregation of omnibus self-ratings of experience in Experiment 1 could have eliminated the contribution of idiosyncratic factors to the benefit of dimensions relevant across several categories.

Discussion

Our survey provides additional information on how subjects in Experiment 1 may have understood and answered our self-rating question about each of the VET categories. First, the dimensions of experience that may be most important or salient for each category appear to vary. While subjects' reports of how long they have shown an interest in a category may be quite important for cars or planes, it is less telling for motorcycles, and while exposure to images may be important for leaves, it may not be as relevant for butterflies. Any approach to the measurement of experience with objects may be complicated by the fact that dimensions of experience may be weighted differently for different categories.

Second, the survey results are consistent with the low correlations between self-ratings of expertise and performance on the VET for each category. When subjects report on their expertise with a category, they are telling us more about their semantic knowledge (i.e., how much they would have to say if they wrote about this category) than about their ability to visually discriminate these objects. Future studies could investigate whether the ratings of visual memory are more correlated with visual ability.

Third, when reporting on their experience with each of the VET categories, subjects in Experiment 1 were to some extent providing us with information that is domain general. Whether they were retrieving domain-general knowledge when answering category-specific questions, we cannot tell; the implications of a domain-general factor for experience with objects remain to be explored. If such a dimension exists, it would be interesting to relate it to personality traits such as conscientiousness, which has been related to motivation to learn (Colquitt & Simmering, 1998).

General discussion

Our results demonstrate the inadequacy of the current standard when it comes to measuring and interpreting the shared variance between object and face recognition. This standard consists of comparing performance with faces with only one category and, in most cases, ignoring experience (Dennett et al., 2012; Wilhelm et al., 2010; Wilmer et al., 2010; Zhu et al.,

	BUT	CAR	LEA	MOT	MUS	OWL	PLA	WAD	Average
Q1. Essay	0.22	0.22	0.20	0.17	0.19	0.11	0.24	0.22	0.20
GEN_average	0.35	0.12	0.19	0.13	0.12	0.17	0.18	0.26	0.19
Q4. Text	0.21	0.27	0.22	0.12	0.16	0.19	0.11	0.05	0.17
Q5. Images	0.13	0.05	0.28	0.15	0.15	0.21	0.00	0.21	0.15
Q6. Duration	0.22	0.27	0.16	−0.01	−0.04	0.19	0.31	0.03	0.14
Q3. Importance	0.11	0.12	0.00	0.27	0.18	0.16	0.07	0.13	0.13
Q2. Visual memory	−0.12	0.05	−0.01	0.08	0.07	−0.02	0.02	0.13	0.03

Table 6. Partial correlations of the ratings on each of the detailed domain-specific questions and the average of the 4 domain-general questions with the self-rating question for each category, controlling for the other 6 predictors. The critical two-tail r-value is 0.14.

2010). Using more than a single object category is critical for putting the magnitude of the relationship between face and object measures in perspective; indeed, we find that performance for different object categories also shares limited variance. Such moderate correlations (~0.3–0.4 between any two categories) seem to simply reflect variance associated with specific categories.

It would seem misguided to try to provide a single number for the shared variance between face and nonface object recognition if this value depends on experience. Over what range of experience should it be estimated? We found that experience with objects predicted neither face nor object recognition but moderated their relationship: Performance in object recognition was increasingly more similar to face recognition with increasing experience. Here, this effect depended on aggregation over several object categories. Future work should explore if the effect can be obtained with single categories given (1) direct measures of domain-general experience, (2) better measures of specific dimensions of category-specific experience such as that which could be estimated using verbal knowledge, or (3) a much larger sample size (Wang et al., submitted).

Our results do not definitively point to a single ability underlying object and face recognition. There could be two distinct abilities that contribute to both CFMT and VET measures, but this appears unlikely because CFMT scores organize around a large primary source of variance (Cho et al., submitted). The more cautious interpretation at this time is that there is no more evidence for a face-specific ability than there is for a butterfly-specific or plane-specific ability. But if the ability to individuate faces and objects reflects a common ability that underlies the acquisition of perceptual knowledge for any visual category, it may be expected to have a common neural substrate. Our results are consistent with high-resolution fMRI results revealing overlap of object and face representations in car and plane experts (McGugin, Gatenby et al., 2012). When expertise is ignored, the face-selective fusiform face area (FFA; Kanwisher, McDermott, & Chun, 1997) shows reliable activity only to faces. But when performance for cars or planes is considered, the location of the expertise effect for cars and planes in the fusiform gyrus cannot be spatially distinguished from face selectivity, even at high resolution.

However, expertise effects in fMRI studies (Bilalic, Langner, Ulrich, & Grodd, 2011; Gauthier et al., 2000; Harley et al., 2009; McGugin, Gatenby et al., 2012; Xu, 2005) rely on *performance* with a category, without consideration for the interaction between experience and ability in accounting for such performance. A meta-analysis of the relationship between car expertise and the selective response to cars in the FFA across all published studies (Gauthier et al., 2000; Gauthier & Curby, 2005; Gauthier, Curby, Skudlarski, & Epstein, 2005; Harel, Gilaie-Dotan, Malach, & Bentin, 2010; McGugin, Gatenby et al., 2012; McGugin, Richler et al., 2012; Van Gulick et al., 2012; Xu, 2005) shows the relationship to be $r = 0.54$, 95% CI [0.40, 0.65], suggesting that only about 25% of the variance is accounted for. Measuring *experience* independently of *performance* may help account for more of the individual differences in brain activity. In addition, experience likely has a larger range than observed here. In many domains that include a perceptual component, experts refine their skills over many years (Ericsson & Lehmann, 1996) up to levels of performance higher than we were able to capture in the current sample. Efforts in sampling a greater range of experience should improve the measurement of object recognition abilities.

Our results may have implications for the interpretation of the high heritability measured for face recognition ability. In Wilmer et al. (2010), the strong correlation for performance on the CFMT in monozygotic twin pairs ($r_{162} = 0.70$) was more than twice that for dizygotic twin pairs ($r_{123} = 0.29$) (see also Zhu et al., 2010), with at least 68% of variability in face recognition explained by genetic factors. Such results have spurred the beginnings of a search for genes that determine the response of the FFA (Brown et al., 2012). The definition and measurement of this behavioral phenotype should be refined to account for the common variance between face and object recognition as a function of experience. Our results also dramatically expand the areas of human activity for which this heritable ability is relevant.

Coda

Ever since we started working with real-world expertise, we have asked people to rate their experience in a domain, only to conclude that such self-ratings have little use since people seem to be unable to predict their *performance* (McGugin, Gatenby et al., 2012; McGugin, Richler et al., 2012). It turns out we should have listened better, as they were providing us with very useful information about their *experience*, and experience is what allows a domain-general ability to become expressed in category-specific performance.

Keywords: individual differences, experience, visual perception

Corresponding author: Isabel Gauthier.
Email: isabel.gauthier@vanderbilt.edu.
Address: Department of Psychology, Vanderbilt University, Nashville, TN, USA.

Footnote

[1]While the data were not collected to test the mediating role of experience, the analyses were performed to test this hypothesis, using the largest available data set at the time.

References

Barton, J. J. S., Hanif, H., & Ashraf, S. (2009). Relating visual to verbal semantic knowledge: The evaluation of object recognition in prosopagnosia. *Brain, 132*(12), 3456–3466, doi:10.1093/brain/awp252.

Bilalic, M., Langner, R., Ulrich, R., & Grodd, W. (2011). Many faces of expertise: Fusiform face area in chess experts and novices. *Journal of Neuroscience, 31*(28), 10206–10214, doi:10.1523/JNEUROSCI.5727-10.2011.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., et al. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and subject-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology, 26*(5), 423–455, doi:10.1080/02643290903343149.

Brown, A. A., Jensen, J., Nikolova, Y. S., Djurovic, S., Agartz, I., Server, A., et al. (2012). Genetic variants affecting the neural processing of human facial expressions: Evidence using a genome-wide functional imaging approach. *Translational Psychiatry, 2*(7), e143, doi:10.1038/tp.2012.67.

Cho, S. J., Wilmer, J., Herzmann, G., McGugin, R. N., Fiset, D., Van Gulick, A. E., & Ryan, K. (submitted). Item response theory analyses of the Cambridge Face Memory Test (CFMT).

Colquitt, J. A., & Simmering, M. J. (1998). Conscientiousness, goal orientation, and motivation to learn during the learning process: A longitudinal study. *Journal of Applied Psychology, 83*(4), 654–665.

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 94–107, doi:10.1037/0096-1523.35.1.94.

Davis, J. M., McKone, E., Dennett, H., O'Connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual differences in the ability to recognise facial identity are associated with social anxiety. *PLoS ONE, 6*(12), e28800, doi:10.1371/journal.pone.0028800.

Dennett, H. W., McKone, E., Edwards, M., & Susilo, T. (2012). Face aftereffects predict individual differences in face recognition ability. *Psychological Science, 23*(11), 1279–1287, doi:10.1177/0956797612446350.

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., et al. (2011). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods, 44*(2), 587–605, doi:10.3758/s13428-011-0160-2.

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 subjects. *Neuropsychologia, 44*(4), 576–585, doi:10.1016/j.neuropsychologia.2005.07.001.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309–331.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273–305.

Gauthier, I., & Curby, K. M. (2005). A perceptual traffic-jam on Highway N170: Interference between face and car expertise. *Current Directions in Psychological Science, 14*(1), 30–33.

Gauthier, I., Curby, K. M., Skudlarski, P., & Epstein, R. (2005). Individual differences in FFA activity suggest independent processing at different spatial scales. *Cognitive, Affective, and Behavioral Neuroscience, 5*(2), 222–234.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6*(4), 428–432, doi:10.1038/nn1029.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits

brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191–197, doi:10.1038/72140.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 431–446.

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition, 118*(2), 201–210, doi:10.1016/j.cognition.2010.11.002.

Harel, A., Gilaie-Dotan, S., Malach, R., & Bentin, S. (2010). Top-down engagement modulates the neural expressions of visual expertise. *Cerebral Cortex, 20*(10), 2304–2318, doi:10.1093/cercor/bhp316.

Harley, E. M., Pope, W. B., Villablanca, J. P., Mumford, J., Suh, R., Mazziotta, J. C., et al. (2009). Engagement of fusiform cortex and disengagement of lateral occipital cortex in the acquisition of radiological expertise. *Cerebral Cortex, 19*(11), 2746–2754, doi:10.1093/cercor/bhp051.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research information support. *Journal of Biomedical Informatics, 42*(2), 377–381.

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.

McGugin, R. W., Gatenby, C., Gore, J., & Gauthier, I. (2012). High-resolution imaging of expertise reveals reliable object selectivity in the FFA related to perceptual performance. *Proceedings of the National Academy of Sciences, USA, 109*(42), 17063–17068.

McGugin, R. W., McKeeff, T. J., Tong, F., & Gauthier, I. (2010). Irrelevant objects of expertise compete with faces during visual search. *Attention, Perception, and Psychophysics, 73*(2), 309–317, doi:10.3758/s13414-010-0006-5.

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, doi:10.1016/j.visres.2012.07.014.

McGugin, R. W., Van Gulick, A. E., Tamber-Rosenau, B. J., Ross, D. A., & Gauthier, I. (in press). Expertise effects in face-selective areas are robust to

clutter and diverted attention, but not to competition. *Cerebral Cortex.*

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94*(1), 18–38, doi:10.1037/0033-2909.94.1.18.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin and Review, 16*(2), 252–257, doi:10.3758/PBR.16.2.252.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72–101.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory and Cognition, 20*(1), 51–68.

Van Gulick, A. E., McGugin, R.W., & Gauthier, I. (2012). Expertise effects for non-face objects in FFA are robust to inattention but not competition. Presented at the 2012 Meeting of the Society for Neuroscience, New Orleans, LA.

Wang, P., & Gauthier, I., & Cottrell, G. W. (2014). Experience matters: Modeling the relationship between face and object recognition. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces— One element of social cognition. *Journal of Personality and Social Psychology, 99*(3), 530–548, doi:10.1037/a0019972.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., et al. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences, USA, 107*(11), 5238–5241.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex, 15*(8), 1234–1242, doi:10.1093/cercor/bhi006.

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science, 9*(2), 111–125, doi:10.1177/1745691613518075.

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., et al. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology, 20*(2), 137–142, doi:10.1016/j.cub.2009.11.067.