Cognition 166 (2017) 42–55

Contents lists available at ScienceDirect

Cognition

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

Original articles

General object recognition is specific: Evidence from novel and familiar objects

Jennifer J. Richler[a], Jeremy B. Wilmer[b], Isabel Gauthier[a,*]

[a] Vanderbilt University, USA
[b] Wellesley College, USA

ARTICLE INFO

Article history:
Received 25 July 2016
Revised 28 April 2017
Accepted 14 May 2017

Keywords:
Object recognition
Face recognition
Individual differences
IQ

ABSTRACT

In tests of object recognition, individual differences typically correlate modestly but nontrivially across familiar categories (e.g. cars, faces, shoes, birds, mushrooms). In theory, these correlations could reflect either global, non-specific mechanisms, such as general intelligence (IQ), or more specific mechanisms. Here, we introduce two separate methods for effectively capturing category-general performance variation, one that uses novel objects and one that uses familiar objects. In each case, we show that category-general performance variance is unrelated to IQ, thereby implicating more specific mechanisms. The first approach examines three newly developed novel object memory tests (NOMTs). We predicted that NOMTs would exhibit more shared, category-general variance than familiar object memory tests (FOMTs) because novel objects, unlike familiar objects, lack category-specific environmental influences (e.g. exposure to car magazines or botany classes). This prediction held, and remarkably, virtually none of the substantial shared variance among NOMTs was explained by IQ. Also, while NOMTs correlated nontrivially with two FOMTs (faces, cars), these correlations were smaller than among NOMTs and no larger than between the face and car tests themselves, suggesting that the category-general variance captured by NOMTs is specific not only relative to IQ, but also, to some degree, relative to both face and car recognition. The second approach averaged performance across multiple FOMTs, which we predicted would increase category-general variance by averaging out category-specific factors. This prediction held, and as with NOMTs, virtually none of the shared variance among FOMTs was explained by IQ. Overall, these results support the existence of object recognition mechanisms that, though category-general, are specific relative to IQ and substantially separable from face and car recognition. They also add sensitive, well-normed NOMTs to the tools available to study object recognition.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Increasingly, an individual differences approach is being used to characterize the mechanisms that underlie cognition. Such an approach can help to clarify the number, real-world relevance, and developmental origins of mechanisms relied upon to complete a given cognitive task (Wilmer, 2008). Here, we use an individual differences approach to better understand the number of separable mechanisms used to recognize objects.

In the study of object recognition, a distinction can be made between domain-specific mechanisms, which are used for a smaller number of object categories (in the extreme, just one), vs. domain-general mechanisms, which are used for a larger number of object categories (in the extreme, all). To date, much of the research on individual differences in object recognition has focused on domain-specificity, and moreover, on the domain-specificity of a single, widely-researched object category: faces (e.g., Duchaine & Nakayama, 2006; Hildebrandt, Wilhelm, Herzmann, & Sommer, 2013; Shakeshaft & Plomin, 2015; Wilhelm et al., 2010; Wilmer et al., 2010, 2012). Here, we take the opposite approach, focusing on domain-generality and aiming to elucidate principles that may apply broadly across a wide variety of object categories.

There are many good reasons to examine domain-general mechanisms, one of which is the potential real-world predictive power of individual differences-based measures. A basic question arises in this context: Can one capture mechanisms that are broad enough to potentially predict behavior across a variety of life situations, yet specific enough to not simply reflect the sorts of highly general mechanisms that are already well-captured by general

* Corresponding author at: Vanderbilt University, PMB 407817, 2301 Vanderbilt Place, Nashville, TN 37240-7817, USA and Department of Psychology, 301 Wilson Hall, Vanderbilt University, Nashville, TN 37240, USA.

E-mail address: isabel.gauthier@vanderbilt.edu (I. Gauthier).

http://dx.doi.org/10.1016/j.cognition.2017.05.019
0010-0277/© 2017 Elsevier B.V. All rights reserved.

intelligence (IQ) tests? Could one, for example, create a test that predicts learning of fingerprints, faces, and X-rays in a group of people who score similarly on IQ tests? Our interest in capturing domain-general components of object recognition is thus driven in part by a desire to identify consequential non-IQ abilities, something that has rarely been achieved in studies of cognitive variation (Schmidt & Hunter, 2004; Wai, Lubinski, & Benbow, 2009).

Our second motivation for focusing on domain-generality is to enhance our understanding of the number of dissociable mechanisms used to recognize objects. Past work in neuropsychology (Farah, 1990, 1992), neuroimaging (Kanwisher, 2000, 2010), individual differences (Dennett et al., 2012; Duchaine & Nakayama, 2006; Wilhelm et al., 2010; Wilmer et al., 2012), and behavioral genetics (Shakeshaft & Plomin, 2015; Wilmer et al., 2010) has frequently focused on a simple dichotomy between faces and objects. This work tends to assume, implicitly or explicitly, that non-face object processing is accomplished via a common set of highly overlapping mechanisms that vary little from one non-face category to another. Not infrequently, this assumption motivates the use of a single non-face object category to test for a dissociation of face processing from domain-general object processing. For example, Shakeshaft and Plomin (2015) concluded, based on results from a face test and a single object test (a car test), that the genes underlying face recognition dissociated from those underlying "general object recognition."

The assumption of common mechanisms for different object categories can, however, be questioned on multiple grounds. First, dissociations have been found between the neural areas supporting the processing of animals vs. tools (e.g., Chao, Weisberg, & Martin, 2002), large vs. small objects (e.g., Konkle & Oliva, 2012) and objects that are curvilinear vs. rectilinear (e.g., Nasr, Echavarria, & Tootell, 2014; Yue, Pourladian, Tootell, & Ungerleider, 2014). Second, behavioral dissociations are found between object categories, and, interestingly, the degree of behavioral dissociation predicts the degree of neural dissociation (Cohen, Konkle, Rhee, Nakayama, & Alvarez, 2014; Cohen, Nakayama, Konkle, Stantić, & Alvarez, 2015). Third, and most relevant to the current focus on individual differences, are recent studies of correlations in performance across object recognition tests (e.g. butterflies, cars, planes, shoes, dinosaurs; McGugin, Richler, Herzmann, Speegle, & Gauthier, 2012; Van Gulick, McGugin, & Gauthier, 2015). The mean pairwise correlation found among these tests ($r = 0.33–0.34$) was no larger than what is typically found between face and non-face object recognition tests (e.g. $r = 0.37$ in Dennett et al., 2012), a result difficult to reconcile with the notion that a single test could capture domain-general object recognition. Moreover, individual pairwise correlations varied widely by category-pair (from $r = 0.00$ for cars and leaves to $r = 0.54$ for leaves and butterflies), suggesting that the contributions of domain-general mechanisms to everyday object recognition may differ sharply from one category to another (McGugin, Richler, et al., 2012; Van Gulick et al., 2015).

Indeed, one might ask whether domain-general mechanisms necessarily contribute at all to individual differences in object recognition. In theory, the modest associations found between object recognition tests might have nothing to do with object recognition *per se*, but might instead reflect more general differences in IQ, attentiveness, or motivation. A key aim of the present work was to verify whether any individual differences in domain-general object recognition exist. A second, related aim was to ask whether individual differences in domain-general object recognition are underestimated by correlations among familiar object categories. In theory, dissociations in performance between object categories could result not only from domain-specific object recognition mechanisms, but also from domain-specific non-perceptual knowledge (e.g. names of car makes and models) gained through domain-specific experience with familiar objects (e.g. extensive research on cars prior to buying one).

Our first two studies test a pair of predictions drawn from the hypothesis that nontrivial individual differences in object recognition exist: (A) measures of object recognition performance that minimize the impact of individual differences in domain-specific experience will correlate relatively highly, via cleaner isolation of domain-general object recognition mechanisms, and (B) associations between such measures will not be substantially explained by measures that are known to load highly on IQ. We tested prediction A in Studies 1 and 2 via different approaches. In Study 1, we created object recognition tests for three novel object categories (Novel Object Memory Tests; NOMTs). The use of novel categories, with which everyone should be similarly unfamiliar, should minimize the impact of individual differences in domain-specific experience. In Study 2, we attempted to minimize the impact of category-specific experience by averaging performance across tests of familiar categories. In both Studies 1 and 2, we then tested prediction B by asking whether controlling statistically for performance on IQ-loaded measures would substantially reduce or eliminate associations between object recognition tests. To preview our results, both predictions held: our efforts to reduce the impact of category-specific experience yielded higher correlations, and these correlations were remarkably impervious to controls for multiple IQ-related measures, thereby supporting the existence of individual differences in domain-general object recognition mechanisms.

Studies 1 and 3 tested two simple predictions of the further hypothesis that the same domain-general mechanisms contribute to recognition of both unfamiliar (novel) and familiar object categories. In Study 1, we examined correlations between novel object recognition and face recognition. Plausibly, recognition in both of these cases may be relatively free of domain-specific experience variation. In the case of face recognition, performance might be relatively free of experience variation if most persons reach a saturation point in their experience whereby only genetic variation remains (this would be consistent with the high heritability found in existing twin studies: Shakeshaft & Plomin, 2015; Wilmer et al., 2010). In the case of novel object recognition, everyone should be similarly inexperienced. If domain-specific experience variation were relatively absent, and if the same underlying mechanisms were used in familiar and unfamiliar object recognition, then performance should correlate highly between faces and novel objects. The correlations we found, however, were weaker than those among NOMTs, tentative evidence that recognition of familiar versus unfamiliar object categories may rely on at least partially distinct mechanisms. In Study 3, we asked whether the NOMTs' relatively low correlation with face recognition is unique to faces, or whether similar results can be obtained using cars, a category that in past work has shown a degree of dissociation from other object categories that is similar to that for faces (McGugin, Richler, et al., 2012; Van Gulick et al., 2015). Cars provide an interesting test case. On the one hand, car recognition is as heritable as face recognition (Shakeshaft & Plomin, 2015), potentially motivating an experience-saturation hypothesis similar to the one mentioned above for face recognition. On the other hand, car recognition is highly correlated with both self-reported car experience and objectively assessed, car-related semantic knowledge, suggesting that statistical controls for one or both might isolate a relatively pure object recognition capacity. Again, however, the correlation of car recognition with NOMTs was weaker than those among NOMTs, even after controlling for experience and semantic knowledge, further evidence that recognition of familiar versus unfamiliar object categories may rely on distinct mechanisms. To summarize, face and car recognition both correlate relatively little with novel object recognition compared with the inter-correlations between NOMTs. This result suggests that domain-general mechanisms underlying familiar object recognition may differ, at least in part, from those underlying unfamiliar object recognition.

The core data analyzed for Studies 1, 2, and 3 are posted as Supplemental information and contribute to a more extensive collection of normative data for the visual recognition and IQ tests investigated here, that is being published in parallel (Wilmer, Richler, Gauthier, & Germine, submitted for publication). The data for these papers is available at the following Open Science Framework links (present paper: osf.io/6c4m7; normative data set paper: osf.io/qygs4).

2. Study 1

2.1. Methods

2.1.1. Participants

Participants were visitors to TestMyBrain.org. Participation in tests advertised on the website is voluntary, no compensation is provided, and participants can quit at any time. A test battery that included our new Novel Object Memory Tests (NOMTs) was advertised as Recognize That Thing, and 1002 participants (435 male, 554 female, 13 not disclosed; mean age = 32.6 years, SD = 13.94, range = 9–88, 4 not disclosed) completed the full battery. A separate battery that included four IQ-related tests was advertised as Puzzles and Words, and 10,000 participants (3927 male, 6073 female; mean age = 32.19, SD = 15.45, range = 10–100) completed that full battery (full IQ data set is available at osf.io/qygs4 and described in Wilmer et al., submitted for publication). A subset of 105 participants (39 male, 64 female, 2 not disclosed; mean age = 32.9 years, SD = 13.7, range = 13–75, 1 not disclosed) completed both batteries.[1]

2.1.2. Recognize that thing test battery

The Recognize That Thing I test battery included a Recognition Questionnaire (see Appendix A) that probed self-reported face recognition (12 questions, e.g., "I find it hard to keep track of characters in TV shows and movies," "Compared to my peers, I think my face recognition skills are ...") and general object recognition (8 questions, e.g., "I can recognize my own baggage at the airport," "How easily do you learn to recognize objects visually?") abilities, followed by two Novel Object Memory Tests (NOMTs, details below), the Cambridge Face Memory Test (CFMT; Duchaine & Nakayama, 2006), a Life Experiences Questionnaire (open-ended questions about occupation and interests), and an SAT Questionnaire (self-report of SAT math and verbal scores). Two of the three NOMTs were randomly selected for each participant and the order of the two NOMT tests was randomly assigned to each participant.

NOMTs were created for three categories of novel objects (Greebles, Ziggerins, and Sheinbugs, see Fig. 1). NOMTs were closely modeled after the CFMT (see Duchaine & Nakayama, 2006). Each NOMT (see Fig. 1) started with a learning phase (trials 1–18), where a target object was shown in three views (3 s per view) followed by three test items where participants had to select which of three objects was the object they had just studied. This was repeated for each of six target objects. In the 54 test phase trials that followed (block 2: trials 19–48; block 3: trials 48–72), participants had to select which of three objects was any of the six studied targets. Targets and distractors were presented from the same view within each test trial. A 20-s study period, with all six targets viewed simultaneously, was provided after the learning phase (block 1) and before the last 24 trials (block 3).

In addition to mirroring the CFMT in the abovementioned ways, the NOMTs mirrored three aspects of the CFMT's trial-by-trial difficulty profile. First, difficulty in the learning phase (trials 1–18) was low (mean performance of 98% for each of the three NOMTs, similar to 97% for the CFMT). The easiness of the learning phase aimed to facilitate active learning of target stimuli, build participant morale, and reinforce understanding of the basic task. Second, difficulty gradually increased over the course of the second block (trials 19–48; correlation of trial number with percentage of participants who answer each trial correctly was −0.41, −0.25, and −0.42 for Greebles, Ziggerins, and Scheinbugs, respectively, similar to −0.30 for CFMT). The relatively easy trials earlier in the second block aimed to avoid a jarring, potentially frustrating effect. Third, difficulty varied substantially over the test phase trials (blocks 2 and 3; SD across trials for percentage of participants who answer the trial correctly was 12%, 12%, and 14% for Greebles, Ziggerins, and Scheinbugs, respectively, similar to 14% for CFMT). Wide variation in difficulty across trials facilitates good discriminability across a range of ability levels (Wilmer et al., 2012). Our approach to achieving varied trial-by-trial difficulty on the NOMTs included efforts to vary the similarity of the foils to other targets (similarity was judged based on the authors' intuitions and was loosely verified in iterations with early versions of the tests).

We made two design choices for the NOMTs that deviated from the CFMT. First, only studied views were tested. In the CFMT, test items show target faces in novel unstudied views after the learning phase (block 1). Second, there were no 'noise trials.' In the CFMT, visual noise is added to face images in the last 24 trials (block 3). Each of these choices aimed to boost NOMT performance into a range that was reasonably comparable to the CFMT, and indeed, average NOMT performance (75%) was comparable to average CFMT performance (76.5% correct). We expected recognition performance for novel objects would naturally tend lower than for faces due to participants' relative lack of experience with the novel objects.

2.1.3. Puzzles and words battery

The Puzzles and Words battery consisted of four IQ-related measures: the TMB Vocabulary test (hereafter Vocabulary) and the TMB Matrices test (hereafter Matrices) – two tests that were developed via the TestMyBrain (TMB) project – plus self-reported SAT verbal (also called "Critical Reading") and SAT math scores.

Vocabulary consisted of 20 items that showed one word printed in capital letters with five response options. On each trial, participants were instructed to select the word that came closest to the meaning of the word printed in capital letters (see Fig. 2 for examples). Measures of vocabulary are among the best indices of verbal or crystallized intelligence and also of general intelligence more broadly (Carroll, 1997). Vocabulary was modeled after the well-validated Wordsum test used in the General Social Survey (Smith, Marsden, Hout, & Kim, 2013). Vocabulary is twice the length of the 10-item Wordsum, and this produces the expected boost in reliability (Cronbach's alpha = 0.84 for Vocabulary in the present sample versus 0.68 for Wordsum; Cor, Haertel, Krosnick, & Malhotra, 2012). Here, Vocabulary correlates robustly with SAT verbal (rho0.50, n = 1358, 95% CIs [0.46, 0.54]); this correlation is comparable to prior reports of correlations between well-validated vocabulary tests and SAT verbal (Mayer & Massa, 2003). As expected (Mayer & Massa, 2003; Rohde & Thompson, 2007), Vocabulary correlates to a lesser degree, but still robustly, with SAT math (rho = 0.29, n = 1345, 95% CIs [0.24, 0.33]) and with Matrices (rho = 0.32, n = 10,000, 95% CIs [0.31, 0.34]).[2]

[1] At the request of a reviewer, we ran an additional battery that looked at NOMT and Digit Span performance (Supplemental info). These data converged with the other results of Study 1, showing only very modest correlations of NOMTs with Digit Span.

[2] Correlations in this and the next two paragraphs are Spearman rho after age is regressed out of Vocabulary and/or Matrices via third-order fit. The use of Spearman's rho minimizes the impact of any outliers. Controlling for age avoids suppression of relationships due to different age curves for Vocabulary and Matrices and due to the absence of an age curve for SAT because SAT is generally taken at a uniform age.

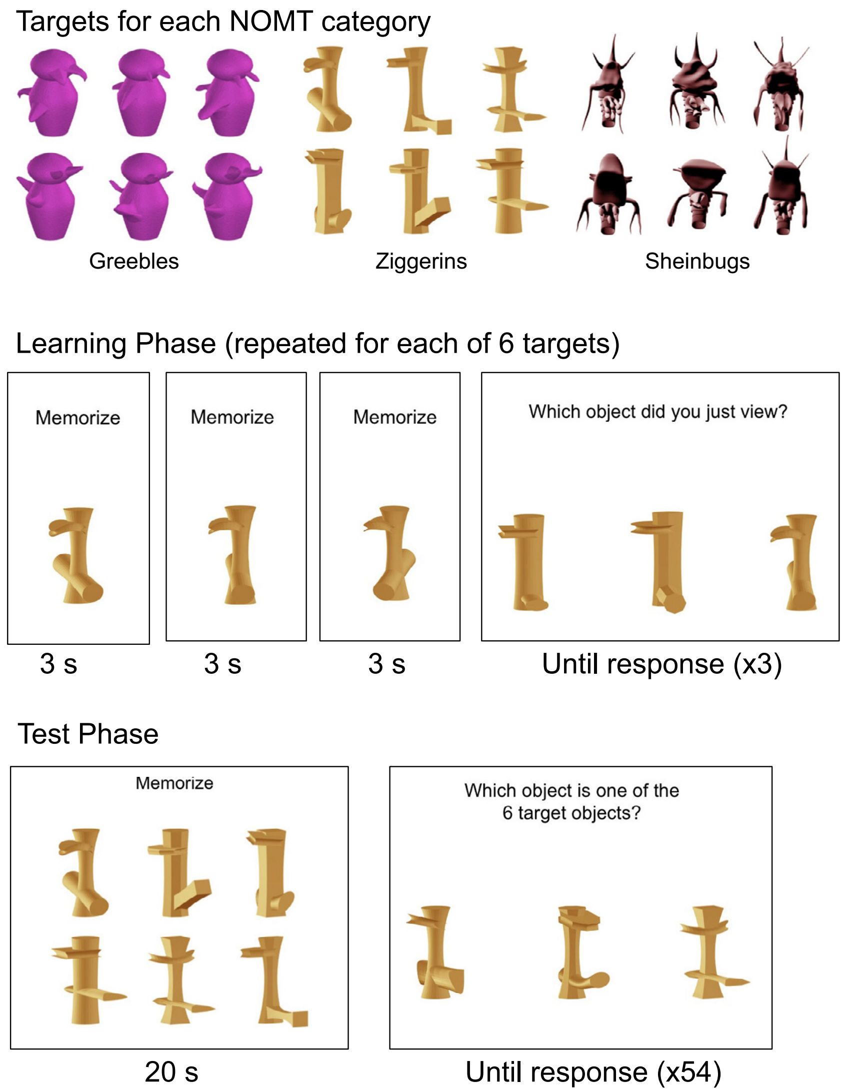

Fig. 1. Top panel: Targets for each of the three NOMTs (Greebles, Ziggerins, and Sheinbugs). Bottom panel: Illustration of NOMT test format (learning and test phase) with Ziggerins.

Matrices consisted of 35 matrix reasoning problems. On each trial, participants are asked to select the image that best completes the pattern (see Fig. 2 for examples). Measures of matrix reasoning are among the best indices of fluid intelligence and also of general intelligence more broadly (Carroll, 1997). Matrices was modeled after the well-validated Matrix Reasoning test used in the Wechsler Abbreviated Scale of Intelligence II (Wechsler & Hsiao-pin, 2011). Matrices has similar reliability to the original WASI II Matrix Reasoning test (Spearman-Brown corrected split-half reliability, computed as in the WASI II manual, which counts all trials after its three-consecutive-incorrect stopping rule as incorrect, is 0.89; Cronbach's alpha is 0.76). Here, Matrices correlates robustly with SAT math (rho = 0.40, n = 1345, 95% CIs [0.36, 0.45]); this correlation is comparable to prior reports of correlations between well-validated matrix reasoning tests and SAT math (Rohde & Thompson, 2007). As expected (Rohde & Thompson, 2007), Matrices correlates to a lesser degree, but still robustly, with SAT verbal (rho = 0.24, n = 1358, 95% CIs [0.19, 0.29]) and Vocabulary (rho = 0.32, n = 10,000, 95% CIs [0.31, 0.34]).

SAT verbal and SAT math scores were self-reported (with an option to skip) and were filtered to include only participants who reported plausible multiple-of-10 SAT verbal scores in the range 200–800. SAT data was thus obtained for about 13% of participants (SAT verbal n = 1358, SAT math n = 1345). The SAT has shown high correlations with multiple measures of general intelligence (Condon & Revelle, 2014; Frey & Detterman, 2004). In keeping with prior web-based research, SAT scores, among those who reported them, showed higher means than, but similar standard deviations to, the SAT normative samples (SAT verbal M = 623, SD = 132; SAT math M = 620, SD = 131; Condon & Revelle, 2014). SAT math correlated robustly with SAT verbal (rho = 0.51, n = 1309, 95% CIs ([0.47, 0.55]); this correlation is comparable to prior reports of correlations between self-reported SAT math and SAT verbal (Mayer & Massa, 2003).

2.2. Results

Data from one participant were discarded for below chance performance on both NOMT tests. Given that we report Spearman rank-order correlations below, which are robust to outliers, correlations change little to none if this participant is added back in.

2.2.1. Relations among NOMTs and between NOMTs and the CFMT

Descriptive statistics and reliability for the NOMTs and CFMT are presented in Table 1. Tests were comparable in summary

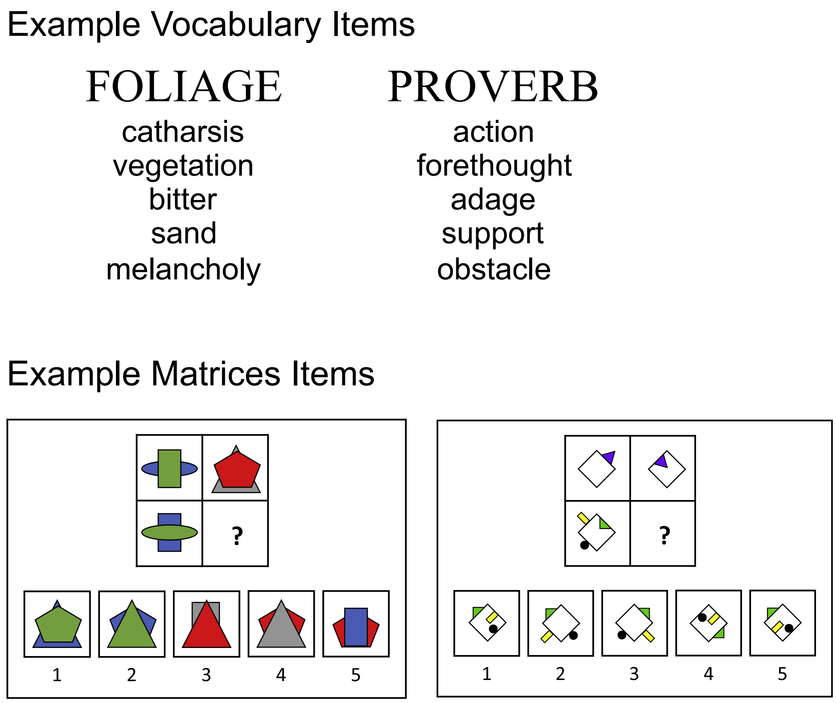

Fig. 2. Example Vocabulary and Matrices items from the Puzzles and Words test battery. In Vocabulary (top panel) participants are instructed to choose the word that best matches the definition of the word in capitals. In Matrices (bottom panel) participants are instructed to choose the item that best completes the pattern. Stimuli shown are different from, but chosen to be representative of, those used in the actual tests.

statistics and performance range, and all tests showed good reliability (Cronbach's $\alpha \geq 0.8$).

Correlations (Spearman's rho) between all NOMTs and the CFMT are reported in Table 2. Because each participant was randomly assigned to complete the CFMT plus two of the three NOMTs, N varies between correlations, but is large (>325) in all cases. Consequently, all correlations greater than 0.1 are statistically significant ($p < 0.001$), and we have therefore opted not to report p-values.

The NOMTs were more strongly correlated with each other (Fisher-transformed average[3] $r = 0.48$, $r^2 = 0.23$) than with the CFMT (average $r = 0.33$, $r^2 = 0.11$), both on average and for each individual pairwise correlation, although the smallest NOMT correlation (between Sheinbugs and Ziggerins, $r = 0.43$) is not significantly different from the largest correlation with the CFMT (between CFMT and Greebles, $r = 0.37$; Fisher's $Z = 1.06$, $p = 0.29$).

An average NOMT z-score was computed for each participant. The correlation between average NOMT and the CFMT was slightly higher (n = 1001, $r = 0.38$, 95% CI = [0.33, 0.43], $r^2 = 0.14$) than correlations between the CFMT and any one NOMT, consistent with known benefits of averaging (Rushton, Brainerd, & Pressley, 1983).

Age and gender were not correlated with CFMT performance (age: n = 997, $r = 0.06$, 95% CI = [−0.002, 0.12], $r^2 = 0.004$; gender: n = 988, $r = -0.03$, 95% CI = [−0.09, 0.03], $r^2 = 0.001$), but they accounted for a small but significant amount of the variance in average NOMT performance (age: n = 997, $r = -0.12$, 95% CI = [−0.18, −0.06], $r^2 = 0.01$; gender coded female = 1, male = 2: n = 988, $r = 0.11$, 95% CI = [0.05, 0.17], $r^2 = 0.01$).

2.2.2. Relations with intelligence measures

Among those who completed the Recognize That Thing test Battery, self-reported SAT-verbal and SAT-math scores correlated robustly (n = 137, $r = 0.71$, 95% CI = [0.62, 0.78], $r^2 = 0.51$, $p < 0.001$). Matrices and Vocabulary scores showed a statistically significant but small correlation (n = 279, $r = 0.25$, 95% CI = [0.14, 0.36], $r^2 = 0.06$, $p < 0.001$). Matrices correlated with SAT-math (n = 37, $r = 0.47$, $r^2 = 0.22$, 95% CI = [0.17, 0.69], $p = 0.002$) but not SAT-verbal in this sample (n = 36, $r = 0.01$, $r^2 = 0.00$, 95% CI = [−0.32, 0.34], $p = 0.95$); Vocabulary correlated with both SAT-math (n = 37, $r = 0.34$, $r^2 = 0.12$, 95% CI = [0.02, 0.60], $p = 0.02$) and SAT-Verbal (n = 36, $r = 0.44$, $r^2 = 0.19$, 95% CI = [0.13, 0.67], $p = 0.004$).

Correlations (Spearman's rho) between IQ-related measures and each NOMT are shown in Table 3. Matrices and SAT scores were significantly correlated with average NOMT performance (Matrices: n = 279, $r = 0.30$, 95% CI = [0.19, 0.40], $r^2 = 0.09$, $p < 0.001$; SAT: n = 137, $r = 0.19$, 95% CI = [0.02, 0.35], $r^2 = 0.04$, $p = 0.03$), while Vocabulary score was not (n = 279, $r = 0.08$, 95% CI = [−0.04, 0.20], $r^2 = 0.01$, $p = 0.18$). CFMT correlated to a small but significant degree with only Vocabulary among the intelligence measures (see Table 3).

The results suggest that intelligence makes a small but significant contribution to NOMT performance. To determine whether intelligence accounts for the variance that is common *between* NOMTs, data were analyzed separately for subsets of participants who completed the same two NOMTs and one or more IQ measures. Of critical interest is the correlation between NOMTs after controlling for IQ-related measures. As shown in Tables 4 and 5, partial correlations changed little relative to the first-order effects, indicating that IQ does not drive correlations between NOMTs.

2.2.3. Relations with self-reported face and object recognition ability

Self-report for face recognition (Cronbach's $\alpha = 0.88$; n = 887) was more reliable than self-report for object recognition (Cronbach's $\alpha = 0.58$; n = 904), possibly because five of the object questions targeted specific object categories (houses, cars, animals, scenes, baggage) for which participants may rely on different sources of information to make their responses, although in principle averaging across these questions should reduce the contribu-

[3] For all subsequent analyses, average correlations were calculated using Fisher-transform.

Table 1
Descriptive statistics and reliability for the NOMTs and CFMT. For all tests, chance is 33% (24 items correct) and perfect performance is 100% (72 items correct).

	NOMTs			CFMT
	Greebles	Ziggerins	Sheinbugs	
N	673	674	655	1001
Mean	70.3%	84.4%	70.4%	76.5%
SD	10.7%	11.2%	10.6%	14.1%
Minimum	36.1%	36.1%	40.3%	36.1%
Maximum	98.6%	100%	98.6%	100%
Cronbach's α	0.80	0.89	0.80	0.91

Table 2
Correlations (Spearman's rho) between NOMTs and the CFMT. All correlations are significant ($p < 0.001$). 95% confidence intervals shown in parentheses.

	Greebles	Ziggerins	Sheinbugs
CFMT	0.37 (0.30, 0.43) n = 673	0.35 (0.28, 0.41) n = 674	0.28 (0.21, 0.35) n = 655
Greebles		0.50 (0.42, 0.58) n = 346	0.50 (0.41, 0.58) n = 327
Ziggerins			0.43 (0.34, 0.51) n = 328

Table 3
Correlations (Spearman's rho) between intelligence measures and the CFMT and NOMTs. 95% confidence intervals are shown in parentheses.

	Puzzles & words		SAT score
	Matrices	Vocabulary	
CFMT	0.08 (−0.04, 0.20) n = 279	*0.13 (0.01, 0.24) n = 279	0.03 (−0.14, 0.20) n = 137
Greebles	**0.34 (0.21, 0.46) n = 186	0.11 (−0.03, 0.25) n = 186	0.18 (−0.03, 0.37) n = 91
Ziggerins	**0.24 (0.10, 0.37) n = 182	0.04 (−0.11, 0.18) n = 182	0.11 (−0.10, 0.31) n = 89
Sheinbugs	**0.24 (0.10, 0.37) n = 190	0.05 (−0.09, 0.19) n = 190	0.17 (−0.03, 0.3) n = 94

* $p < 0.05$.
** $p < 0.002$.

tion of category-specific factors (cf. Gauthier et al., 2014).[4] Self-report scores for faces and objects were moderately correlated ($n = 838$, $r = 0.53$, 95% CI = [0.48, 0.58], $r^2 = 0.28$).

As shown in Table 6, self-report was a better predictor of performance for faces than novel objects, with the largest correlation between self-reported face recognition ability and performance with faces ($r = 0.39$).

2.3. Discussion

In Study 1, we created the first psychometrically reliable tests of recognition for three novel object categories (Novel Object Memory Tests; NOMTs) and found that correlations in performance among these tests (average 23% shared variance) was, though far from perfect, substantially larger than the correlations previously found among familiar object memory tests (FOMTs; average ~10% shared variance; Dennett et al., 2012; McGugin, Richler, et al., 2012; Van Gulick et al., 2015). We also found that Matrices was correlated with NOMT performance (5–8% shared variance), but not with face recognition performance.[5] Importantly, however,

Table 4
Correlations among NOMTs and IQ-related measures for groups of participants who completed the same two NOMTs and also completed the Puzzles & Words Battery. Partial correlations (Spearman's partial rho) are correlations between NOMTs controlling for the IQ-related measure shown in the current row of the table. Matrices and Vocabulary were combined by taking their mean percentile score.

Ziggerins with Greebles (N = 89) rho b/w NOMTs = 0.63	Ziggerins	Greebles	Partial *rho* b/w NOMTs
Matrices	0.29	0.31	0.59
Vocabulary	0.07	0.13	0.63
Matrices and vocabulary combined	0.23	0.28	0.60
Scheinbugs with Greebles (N = 97) rho b/w NOMTs = 0.47	**Sheinbugs**	**Greebles**	**Partial *rho* b/w NOMTs**
Matrices	0.29	0.38	0.41
Vocabulary	0.10	0.09	0.47
Matrices and vocabulary combined	0.27	0.22	0.44
Scheinbugs with Ziggerins (N = 93) rho b/w NOMTs = 0.56	**Sheinbugs**	**Ziggerins**	**Partial *rho* b/w NOMTs**
Matrices	0.21	0.20	0.54
Vocabulary	0.00	−0.03	0.56
Matrices and vocabulary combined	0.14	0.11	0.55

Table 5
Correlations (Spearman's rho) among NOMTs and SAT for groups of participants who completed the same two NOMTs and reported valid (multiple of 10 between 200 and 800) SAT math and SAT verbal scores. Partial correlations (Spearman's partial rho) are correlations between NOMTs controlling for the self-reported SAT score shown in the current row of the table. SAT math and SAT verbal were combined by adding them together.

Ziggerins with Greebles (N = 48) rho b/w NOMTs = 0.49	Ziggerins	Greebles	Partial *rho* b/w NOMTs
SAT math	0.10	0.13	0.48
SAT verbal	0.04	−0.04	0.49
SAT combined	0.09	0.05	0.49
Scheinbugs with Greebles (N = 43) rho b/w NOMTs = 0.52	**Sheinbugs**	**Greebles**	**Partial *rho* b/w NOMTs**
SAT math	0.45	0.38	0.43
SAT verbal	0.26	0.10	0.52
SAT combined	0.39	0.22	0.49
Scheinbugs with Ziggerins (N = 46) rho b/w NOMTs = 0.36	**Sheinbugs**	**Ziggerins**	**Partial *rho* b/w NOMTs**
SAT math	−0.02	0.24	0.38
SAT verbal	−0.06	0.03	0.37
SAT combined	0.00	0.16	0.37

* $p < 0.05$.
** $p = 0.01$.

none of our IQ-related measures accounted for the correlations between NOMTs. In other words, although intelligence contributed somewhat to performance with novel objects, it did not account for relations in performance between novel object categories. Together, these results suggest that learning and recognizing novel objects is a reliable ability that is separate from intelligence.

[4] In prior work where self-report consisted of the same unique question for all categories, test-retest reliability was much lower for faces than non-face categories (Gauthier et al., 2014).

[5] Other studies have found a modest but significant correlations of fluid intelligence with CFMT ($r = 0.14$, Van Gulick et al., 2015; $r = 0.16$, Shakeshaft & Plomin, 2015).

Table 6
Correlations (Spearman's rho) between self-report face and object recognition ability measures and CFMT and NOMT performance. All correlations are statistically significant ($p < 0.01$) unless otherwise noted. 95% confidence intervals are shown in parentheses.

	Self-report: faces	Self-report: objects
CFMT	0.39 (0.33, 0.44) n = 887	0.22 (0.16, 0.28) n = 904
NOMT average	0.13 (0.07, 0.19) n = 887	0.14 (0.08, 0.20) n = 904
Greebles	0.19 (0.11, 0.27) n = 595	0.15 (0.07, 0.23) n = 597
Ziggerins	0.11 (0.03, 0.19) n = 608	0.12 (0.04, 0.20) n = 619
Sheinbugs	0.05[a] (−0.03, 0.13) n = 571	0.11 (0.03, 0.19) n = 592

[a] Not significant.

The NOMTs were also correlated with the CFMT (11% shared variance on average). Thus, the recognition of faces and novel objects are as different as two familiar categories of non-face objects. In Study 3, we will show that the same level of dissociation is found between cars and novel objects, adding to a number of intriguing parallels found between faces and cars. Together, the relatively weak correlations of cars and faces with NOMTs, relative to the correlations among NOMTs, suggests that the category-general mechanisms captured by NOMTs are not necessarily highly involved in either car recognition or face recognition.

3. Study 2: Reanalysis of Van Gulick et al. (2015) Study 2

In Study 1, we found that while intelligence correlated somewhat with performance with novel objects, it did not account for the shared variance among different novel object categories. In Study 2, we test the generality of this finding in a reanalysis of data from Study 2 of Van Gulick et al. (2015) that included the Vanderbilt Expertise Test for 8 familiar object categories, including cars, as well as a three-task estimate of fluid intelligence.

3.1. Methods

Study 2 of Van Gulick et al. (2015) tested 213 participants (86 male; mean age = 22.49, SD = 6.31, age range: 8–55). We analyzed the original data from the Vanderbilt Expertise Test (VET) for 8 categories (birds, cars, dinosaurs, leaves, mushrooms, planes, shoes and transformers) and the mean score across three IQ-related tests (Raven's Advanced Progressive Matrices, Raven, Raven, & Court, 1998; Letter sets, Ekstrom, French, Harman, & Dermen, 1976; Number series, Thurstone, 1938). Details of the tasks, reliability (0.71–0.92 for the VET tasks, 0.92 for IQ), and correlations with gender and age can be found in the original paper. Although Pearson's r was reported in the original paper, we used Spearman's rho here to be consistent with Studies 1 and 2. The VET tasks were, in certain key ways, modeled after the CFMT (Duchaine & Nakayama, 2006) and Cambridge Car Memory Test (CCMT; Dennett et al., 2012): in each VET, six targets are studied, followed by 48 3-AFC trials increasing in difficulty. Key differences are: (1) the VETs exclude the block of 18 learning trials at the beginning of CFMT and CCMT, and while the CFMT and CCMT then allow a 30 s study period for all six faces, the VETs begin by allowing participants to study all six target objects for a duration of their choice, and (2) CFMT and CCMT images have blank backgrounds, whereas VET images include backgrounds.

3.2. Results

Pairwise correlations between any two VET categories ranged from 0.08 to 0.47 (average = 0.30; see Table 7). As in Study 1, we found that averaging performance across categories is beneficial, in this case likely because noise, experience, and/or interest are somewhat independent between VET categories. Table 8 shows two ways of calculating a grand average representing correlations across different familiar categories. The first and most intuitive is the average of all pairwise Z-transformed correlations ($r = 0.32$). The second approach first correlates performance for each category with the aggregate (average) z-scores of the other 7 categories, before averaging these correlations across categories. This second approach uses the well-known "principle of aggregation" which states that "the sum of a set of measurements is a more stable and representative estimator then any single measurement" (Rushton et al., 1983) and it provides a higher estimate of domain-general variance ($r = 0.49$), or an r^2 (0.24) comparable to the shared variance observed between NOMTs in Study 1. On average, VETs shared a small amount of significant variance with IQ ($r^2 = 0.04$). Most critically, controlling for IQ did not affect the correlations between each category and the average of all 7 other categories (see Table 8; average $r^2 = 0.22$).

3.3. Discussion

This re-analysis reveals a pattern of results for familiar objects that is strikingly consistent with what we observed for NOMTs in Study 1. First, we found evidence for 24% shared variance across familiar categories (vs. 23% for novel objects). The similarity of these two numbers is not important *per se*, as these analyses differed in multiple ways. An important parallel, however, is that in both cases, substantially more variance is explained than by pairwise correlations between familiar object memory tests (10%). This substantial shared variance raises the possibility that averaged FOMTs effectively capture domain-general object recognition mechanisms by averaging out effects of category-specific factors such as experience, interest, and visual similarity. An alternative explanation, however, is that the shared variance reflects more global and non-specific mechanisms such as IQ. By showing that IQ did not contribute an appreciable amount to the shared variance between familiar categories, we provide clear evidence against that alternative explanation. We conclude that averaging across categories reveals mechanisms that are both domain-general and specific relative to IQ.

4. Study 3

In Study 3, we sought to weave a second familiar object category, cars, into our investigation. We chose cars for several reasons. First, cars are often used as a comparison to faces (e.g., McGugin, Gatenby, Gore, & Gauthier, 2012; Shakeshaft & Plomin, 2015), and sometimes assumed to be a representative non-face object category (e.g., Dennett et al., 2012). Second, previous work found that performance with cars tends to dissociate at least as much as face recognition does from performance with other familiar object categories (McGugin, Richler, et al., 2012; Van Gulick et al., 2015), although the correlation of car recognition with face recognition is typical of other non-face categories (Dennett et al., 2012; McGugin, Richler, et al., 2012; Van Gulick et al., 2015). Car recognition was also recently found to be as heritable as face recognition, and each was only modestly correlated with g (Shakeshaft & Plomin, 2015). Therefore, car recognition is similar to face recognition in its heritability and dissociation from other object categories. Here, we test whether car recognition will also dissociate from NOMTs, just as face recognition did in Study 1. We also expect to replicate the finding that the correlation of car with face recognition is neither unusually high, nor unusually low, compared to what is found for other categories (Dennett et al., 2012;

Table 7
Correlations (Spearman's rho) between the different VET categories in Study 2 of Van Gulick et al. (2015).

	Car	Bird	Dino	Leaf	Mush	Plane	Shoe
Bird	0.08						
Dino	0.11	0.41					
Leaf	0.23	0.39	0.34				
Mush	0.06	0.31	0.23	0.43			
Plane	0.24	0.37	0.45	0.46	0.35		
Shoe	0.21	0.22	0.22	0.39	0.24	0.29	
Trans	0.23	0.44	0.34	0.46	0.38	0.47	0.26

Table 8
For each VET category, Z-transformed average pairwise correlations with all other categories, correlation with the average of the other categories, correlation with IQ, and partial correlation with the average of the other categories controlling for IQ.

Category	*r*			Partial *r*
	Average pairwise with other 7 categories	Average of other 7 categories	IQ	Average of other 7 categories
Car	0.17	0.25	0.00	0.28
Bird	0.32	0.50	0.31	0.44
Dino	0.30	0.45	0.26	0.43
Leaf	0.37	0.63	0.21	0.60
Mush	0.28	0.46	0.20	0.42
Plane	0.36	0.61	0.28	0.59
Shoe	0.25	0.40	0.07	0.41
Trans	0.35	0.59	0.27	0.54
Avg Z-transformed *r*	0.32	0.49	0.20	0.47

McGugin, Richler, et al., 2012; Van Gulick et al., 2015). To foreshadow our results, we find that car and face recognition dissociate to about an equal degree from NOMTs, suggesting that the category-general mechanisms that appear to contribute strongly to NOMT performance make a similarly limited contribution to both car and face recognition.

We also measured self-reported experience with cars to test the hypothesis that variation in experience may be a suppressor variable in the relation between car recognition performance and NOMTs (for which there is no variation in experience). We also measured semantic knowledge for cars (knowledge of car models), another method used in prior work to estimate experience with cars (Van Gulick et al., 2015). In that work, for most object categories tested, self-report and semantic knowledge accounted for common portions of the variance in recognition performance. For cars, however, the correlation between car recognition and car semantic knowledge remained high even after partialing out self-reported experience. Interestingly, this was the case even though self-reported experience was a better predictor of car recognition ($r = 0.42$) than for most other categories. In other words, for cars, self-report and semantic knowledge were both related to car recognition performance and appeared to index different aspects of experience. We thus use them both here to test our conjecture that experience could explain why car recognition would dissociate from novel object recognition.

4.1. Methods

4.1.1. Participants

Participants were visitors to TestMyBrain.org. One thousand and eight participants (432 male, 568 female, 8 not disclosed; mean age = 35.5 years, SD = 13.66, range = 7–100, 4 not disclosed) completed a new Recognize That Thing battery. A subset of 160 participants (56 male, 103 female, 1 not disclosed; mean age = 35.2 years, SD = 14.33, range = 7–69) also completed the Puzzles and Words test battery. One thousand and three different participants (343 male, 643 female, 1 not disclosed; mean age = 35.7 years, SD = 14.93, range 9–85) completed a Remembering That Face battery (full data set from this battery is available at osf.io/qygs4 and described in Wilmer et al., submitted for publication).

4.1.2. Batteries

The Recognize That Thing II battery included a Recognition Questionnaire (same as Study 1, see Appendix A), a Car Recognition Questionnaire (7 items that probe self-reported car recognition ability and experience; see Appendix B), one NOMT (randomly chosen for each participant), the Cambridge Car Memory Test (CCMT; Dennett et al., 2012), the SAT questionnaire, a Semantic Vanderbilt Expertise Test for cars (adapted from Van Gulick et al., 2015), and a Life Experience Questionnaire (same as Study 1).

Remembering That Face included the CFMT, the CCMT, and a Life Experience Questionnaire (same as Study 1).

The Cambridge Car Memory test is identical in format to the CFMT, but uses cars instead of faces (see Dennett et al., 2012, for a complete description).

The Semantic Vanderbilt Expertise Tests (SVETs; Van Gulick et al., 2015) measure non-visual semantic category-specific knowledge as estimated through knowledge of domain-relevant nomenclature. Three object names (one real name and two foils) are presented on each trial, and participants are instructed to select the real name. The SVETs have good reliability and domain-specific validity (Van Gulick et al., 2015). Here, we used a shorter 12-item version of the SVET-car, with items selected on the basis of an item analysis of the Van Gulick et al. (2015) data.

4.2. Results

4.2.1. Relations between NOMTs and the CCMT

Descriptive statistics and reliability for the NOMTs and CCMT are presented in Table 9. Tests were comparable in summary statistics and performance range, and all tests showed good reliability (Cronbach's $\alpha \geq 0.78$).

Correlations (Spearman's rho) between each NOMT and the CCMT are reported in Table 10. For ease of comparison, Table 10 also shows correlations between each NOMT and the CFMT from Study 1. Overall, performance with different novel object categories is equally correlated with performance for faces (Study 1,

Table 9
Descriptive statistics and reliability for the NOMTs and CCMT. For all tests, chance is 24 and perfect performance is 72.

	NOMTs			CCMT
	Greebles	Ziggerins	Sheinbugs	
N	336	363	309	1008
Mean	50.43	62.25	50.79	54.38
SD	7.31	8.06	7.45	9.69
Minimum	26	29	23	25
Maximum	68	72	70	72
Cronbach's α	0.78	0.90	0.81	0.88

Table 10
Correlations (Spearman's rho) between each NOMT and the CCMT (Study 2) and CFMT (Study 1). All correlations are significant (p < 0.001). Results of statistical tests comparing correlations with CCMT and CFMT for each NOMT are also shown. 95% confidence intervals are shown in parentheses.

	CCMT	CFMT	*z*	*p*
Greebles	0.40 (0.31, 0.49) n = 336	0.37 (0.30, 0.43) n = 673	0.45	0.65
Ziggerins	0.34 (0.25, 0.43) n = 363	0.35 (0.28, 0.41) n = 674	0.24	0.81
Sheinbugs	0.22 (0.11, 0.32) n = 309	0.28 (0.21, 0.35) n = 655	0.83	0.41

Table 11
Descriptive statistics and reliability for the CCMT and CFMT. For both tests, chance is 24 and perfect performance is 72.

	CCMT	CFMT
N	1003	1003
Mean	49.79	51.28
SD	9.48	10.48
Minimum	22	25
Maximum	72	72
Cronbach's α	0.88	0.90

average $r = 0.33$, $r^2 = 0.11$) and cars (Study 2, average $r = 0.32$, $r^2 = 0.10$), and these correlations are smaller than those obtained between different novel object categories (Study 1, average $r = 0.48$, $r^2 = 0.23$). Notably, the order in which NOMTs account for CCMT performance is the same as that in which they accounted for CFMT performance (Greebles > Ziggerins > Sheinbugs). Indeed, despite considerable statistical power, correlations between each NOMT and the CCMT are not statistically different from the correlations between each NOMT and the CFMT (*zs* < 1, *ps* > 0.4; see Table 10). This is inconsistent with the idea that visual similarity primarily drives these effects.

Replicating Study 1, the average correlation between NOMT performance and age was small but significant (average $r = -0.15$, $r^2 = 0.02$). On average, there was no correlation between NOMT performance and gender (average $r = 0.04$, $r^2 = 0.002$). The correlation between age and CCMT performance was statistically significant (n = 1004, $r = 0.07$, 95% CI = [0.01, 0.13], $r^2 = 0.004$), but accounts for less than 1% of the variance. Gender was significantly correlated with CCMT performance (n = 1000, $r = 0.34$, 95% CI = [0.28, 0.39], $r^2 = 0.12$), replicating a car-advantage in men (Dennett et al., 2012; McGugin, Richler, et al., 2012).

4.2.2. Relations between the CCMT and the CFMT

Descriptive statistics and reliability for the CCMT and the CFMT for this study are presented in Table 11. Mean performance of both tests in this battery was somewhat lower than in our other data (see Tables 1 and 9), but both tests showed good reliability (Cronbach's $\alpha \geq 0.88$). The Spearman correlation of the CCMT with the CFMT was 0.35 (n = 1003, 95% CI = [0.29, 0.40], $r^2 = 0.12$).

4.2.3. Relations with IQ measures

As in Study 1, self-reported SAT-verbal and SAT-math scores were correlated (n = 107, $r = 0.57$, 95% CI = [0.43, 0.69], $r^2 = 0.32$, $p < 0.001$). Matrices and Vocabulary scores were not correlated (n = 158, $r = -0.008$, 95% CI = [−0.15, 0.16], $r^2 < 0.001$, $p > 0.9$). Only 25 participants completed Puzzles & Words and provided SAT scores, so the relationship between these measures is not considered.

Correlations (Spearman's rho) between intelligence measures and each NOMT are shown in Table 12 and generally replicate Study 1: on average Matrices performance accounted for 7% of the variance in NOMT performance (average $r = 0.26$), but Vocabulary and SAT scores were not significant predictors. Matrices was also a significant predictor of CCMT performance (3% shared variance), but Vocabulary and SAT scores were not (see Table 12).

4.2.4. Relations with self-report measures & SVET-car

We used a subset of three items from the Recognition Questionnaire (see Appendix A) that probed general recognition ability (rather than ability with specific categories) as an index of self-reported general visual ability. All self-report measures showed good reliability (Cronbach's α; cars = 0.93, n = 937; faces = 0.88, n = 999; general visual ability = 0.72, n = 981). Reliability for the 12-item SVET-car was moderate (Cronbach's α = 0.65; n = 999).

Correlations between self-report measures and the SVET-car are shown in Table 13, and correlations between these measures and NOMT and CCMT performance are shown in Table 14. Self-report measures were modestly correlated with each other (average $r = 0.28$, $r^2 = 0.08$). Although all self-report measures were significantly correlated with SVET-car, self-report for cars accounts for substantially more variance in performance (20% vs. ~1%). Self-reported car recognition and SVET-car were significant predictors of CCMT performance (20% and 25% variance explained, respectively), replicating similar work using the VET-car instead of the CCMT (Van Gulick et al., 2015). Other correlations between NOMT or CCMT performance and self-report or SVET-car are statistically significant, but account for substantially less variance (1–3%; see Table 14).

4.2.5. Testing a simple experience account

It may be that experience with familiar object categories reduces their correlation with novel objects, with which partici-

Table 12
Correlations (Spearman's rho) between intelligence measures and the CCMT and NOMTs. 95% confidence intervals are shown in parentheses.

	Puzzles & words		SAT SCORE	
	Matrices	Vocabulary	Math	Verbal
CCMT	0.18* (0.03, 0.33) n = 160	0.11 (−0.05, 0.26) n = 158	0.16 (−0.03, 0.34) n = 108	0.11 (−0.08, 0.29) n = 112
Greebles	0.28* (0.01, 0.51) n = 52	−0.01 (−0.28, 0.27) n = 51	0.12 (−0.22, 0.43) n = 36	0.29 (−0.03, 0.55) n = 39
Ziggerins	0.27* (0.02, 0.49) n = 60	−0.14 (−0.38, 0.12) n = 59	0.25 (−0.07, 0.52) n = 40	0.21 (−0.10, 0.49) n = 41
Sheinbugs	0.23 (−0.06, 0.48) n = 48	−0.22 (−0.27, 0.07) n = 48	0.13 (−0.23, 0.46) n = 32	0.19 (−0.17, 0.51) n = 32

* $p < 0.05$.

Table 13
Correlations (Spearman's rho) between self-report measures and SVET-car. All correlations are significant at $p < 0.001$ unless noted. 95% confidence intervals are shown in parentheses.

	Self-report general	Self-report cars	SVET
Self-report faces	0.31 (0.25, 0.37) n = 981	0.21 (0.15, 0.27) n = 937	0.08[a] (0.02, 0.14) n = 985
Self-report general		0.32 (0.26, 0.38) n = 937	0.12 (0.06, 0.18) n = 981
Self-report cars			0.45 (0.40, 0.50) n = 985

[a] $p = 0.01$.

Table 15
Correlations (Spearman's rho) between the CCMT and each NOMT, and partial correlations controlling for self-reported car ability or SVET-car. 95% confidence intervals are shown in parentheses. All correlations are significant ($p < 0.001$).

	r	Partial *r*	
		Self-report	SVET-car
Greebles	0.40 (0.31, 0.49) n = 336	0.38 (0.28, 0.47) n = 309	0.39 (0.30, 0.48) n = 333
Ziggerins	0.34 (0.25, 0.43) n = 363	0.38 (0.29, 0.47) n = 340	0.33 (0.24, 0.42) n = 359
Sheinbugs	0.22 (0.11, 0.32) n = 309	0.31 (0.20, 0.41) n = 288	0.24 (0.13, 0.34) n = 307
Average Z-transformed *r*	0.33	0.37	0.33

pants do not vary in experience. To test whether experience with cars may be suppressing the magnitude of the correlation between cars and NOMTs, we used the Car Recognition Questionnaire (see Appendix B) and the SVET-car as measures of experience (see Gauthier et al., 2014; Van Gulick et al., 2015). As shown in Table 15, partialing out self-reported car recognition or SVET-car does not change the correlations between each NOMT and the CCMT (difference between r and partial r, all $z < 1.2$, $p > 0.2$). Thus, these aspects of experience with cars do not explain smaller correlations between NOMTs and the CCMT compared to NOMTs with each other. A similar analysis using the self-report measure for faces in Experiment 1 yields the same result: Self-reported face recognition ability does not modulate the relationship between the CFMT and NOMTs (correlation between CFMT and NOMT-average after partialing out self-report for faces is $r = 0.35$, from 0.38 for the zero-order correlation).

It is also possible that experience could explain the larger correlations with fluid intelligence for novel objects than with cars. For example, experience may reduce reliance on abstract reasoning or intelligence (perhaps because the category space is well-defined in those with sufficient experience). As shown in Table 16, correlations of CCMT with Matrices and Vocabulary showed modest evidence of suppression by experience, while correlations of CCMT with SAT math and SAT verbal showed modest evidence of the opposite.

Table 16
Correlations (Spearman's rho) between the CCMT and intelligence measures, and partial correlations controlling for self-reported experience or SVET-Car. 95% confidence intervals are shown in parentheses.

	r	Partial *r*	
		Self-report	SVET-car
Matrices	18* (0.03, 0.33) n = 160	0.14 (−0.02, 0.29) n = 157	0.25** (0.10, 0.39) n = 159
Vocabulary	0.11 (−0.05, 0.26) n = 158	0.17* (0.01, 0.32) n = 155	0.25** (0.10, 0.39) n = 159
SAT math	0.16 (−0.03, 0.34) n = 108	0.12 (−0.08, 0.31) n = 103	−0.01 (−0.17, 0.15) n = 157
SAT verbal	0.11 (−0.08, 0.29) n = 112	0.09 (−0.10, 0.28) n = 107	0.05 (−0.14, 0.23) n = 111

* $p < 0.05$.
** $p < 0.001$.

Table 14
Correlations (Spearman's rho) between self-report measures and SVET-car and CCMT and NOMT performance. 95% confidence intervals are shown in parentheses.

	Self-report			
	Faces	General	Cars	SVET-car
CCMT	0.18** (0.12, 0.24) n = 985	0.18** (0.12, 0.24) n = 981	0.51** (0.46, 0.56) n = 937	0.45** (0.40, 0.50) n = 999
Greebles	0.08 (−0.03, 0.19) n = 328	0.08 (−0.03, 0.19) n = 302	0.18* (0.07, 0.29) n = 309	0.12* (0.01, 0.22) n = 333
Ziggerins	0.12* (0.02, 0.22) n = 355	0.08 (−0.02, 0.18) n = 353	0.01 (−0.10, 0.12) n = 340	0.10* (−0.003, 0.20) n = 359
Sheinbugs	0.17* (0.06, 0.28) n = 302	0.12* (0.008, 0.23) n = 302	−0.13* (−0.24, −0.02) n = 288	0.02 (−0.09, 0.13) n = 307

* $p < 0.05$.
** $p < 0.001$.

4.3. Discussion

Study 3 revealed similar correlations between the CCMT and NOMTs and between CCMT and CFMT as were found between the CFMT and NOMTs in Study 1. All of these correlations were smaller than those observed among novel object categories in Study 1. Together, these results suggest that category-specific experience either reduces reliance on category-general object processing mechanisms and/or increases reliance on category-specific mechanisms. The similarly small correlations found for both cars and faces challenge a simple account whereby all non-face object recognition strongly tap a singular domain-general object processing ability. Moreover, correlations between each NOMT and the CCMT and CFMT were highly similar, which is inconsistent with a role for visual similarity in explaining these effects. We also ruled out a simple experience explanation: regressing out experience measures did not influence the correlations between the CCMT or CFMT and NOMTs, suggesting that correlations between familiar and novel object categories are not mediated by the kinds of experience captured by our measures.

5. General discussion

We developed three new reliable tests of novel object recognition (Novel Object Memory Tests; NOMTs). Correlations among NOMTs were substantially stronger than correlations found in this study and, on average, in past work, among tests of familiar object recognition (Familiar Object Memory Tests; FOMTs; McGugin, Richler, et al., 2012; Van Gulick et al., 2015). Critically, these differences are not confounded by differential test reliabilities. Moreover, the shared variance among NOMTs remained essentially identical when controlling for performance on multiple IQ-related tests, ruling out a contribution of a host of global, non-specific mechanisms captured by such tests. We conclude that the substantial variance shared among NOMTs reflects mechanisms that are simultaneously category-general, yet specific relative to IQ. We obtained remarkably similar results with FOMTs. Averaging across FOMTs substantially increased the variance explained in other single FOMTs, relative to pairwise correlations across FOMTs. Again, these differences were above and beyond what could be expected by enhanced test reliabilities. And again, controlling for IQ-related tests caused virtually no reduction in variance explained. These results with FOMTs thus support a highly similar inference to the one obtained from NOMTs: that of mechanisms that are both category-general and yet specific relative to IQ.

The shared variance between NOMTs, as well as that between familiar object categories, provides strong evidence for the existence of domain-general mechanisms that contribute to object recognition. However, one question is whether the same domain-general mechanisms are used across the full spectrum of category-specific familiarity that runs from novel objects at the low end to cars and faces at the high end. Our finding that correlations between NOMTs and the CFMT, or between NOMTs and the CCMT, were smaller than correlations among NOMTS, and no larger than the correlation of CFMT with CCMT, suggests that familiar categories may engage mechanisms that are not used for novel objects (e.g., long-term visual representations or non-visual knowledge). Admittedly, none of our measures of experience were found to mediate the relations between faces/cars and novel objects, even though the experience measures themselves demonstrate some validity by predicting performance. It may be that experience matters, but does not influence performance in as fine-grained a manner as tested here. That is, we regressed out variability in car experience, but certainly all of our participants have experience with cars. Indeed, in past work with multiple object categories, cars were rated, on average, second only to faces in the degree to which they had been experienced. In contrast, participants were seeing our novel objects for the first time. This difference in overall level of experience may, in and of itself, differentiate familiar from novel object recognition. Alternatively, besides semantic knowledge and the self-report measures obtained, there may be other aspects of experience that could account for dissociations with novel objects. For instance, recent work found that gregariousness (Li et al., 2010) and hometown size (Balas & Saville, 2015) are related to face recognition ability. Clearly, such correlational results are causally ambiguous. That is, it could be that over generations, a family whose genes support exceptional face recognition develop a bit of extra gregariousness and a relative comfort with the social complexities of moving to a larger town. Yet such results also raise the possibility that individual differences in social experience, *per se*, cause individual differences in face recognition ability. If so, then prior assertions that face recognition is saturated in experience throughout the normal population (e.g., Gauthier et al., 2014) may have been overstated. Although we found no evidence that experience is suppressing the correlations between CFMT/CCMT and NOMTs, we acknowledge that the measurement of various aspects of experience is still in its infancy.

Alternatively, faces and cars may be unusual not only in their high degree of familiarity, but also in other ways that cause them to engage different mechanisms. Faces and cars have shown several intriguing parallels in past work. First, we found that self-report was a good predictor of performance with faces and cars, an effect that has been found before for faces, cars, and, in one study, shoes (Van Gulick et al., 2015), but not other familiar object categories (McGugin, Richler, et al., 2012; see also Barton, Hanif, & Ashraf, 2009; McGugin, Gatenby, et al., 2012). Second, performance with faces and cars dissociates from performance with other familiar object categories (McGugin, Richler, et al., 2012; Van Gulick et al., 2015). Third, both face and car recognition abilities are substantially and similarly heritable (Shakeshaft & Plomin, 2015; Wilmer et al., 2010). Our results do not indicate why people seem exceptionally good (at least relative to other categories) at predicting their performance for faces and cars, or whether this can help explain their lower correlation with domain-general visual ability. Ultimately, the more categories we can test that vary systematically both in their familiarity and in other ways, the more conclusively we can determine the number of domain-general abilities that contribute to object recognition.

Indeed, our results reveal benefits of testing multiple object categories for interpreting specific correlations. For example, the similarity in the magnitude of the correlations between cars and faces with each novel object category is striking, and could suggest systematic relationships between our specific novel object categories and familiar object categories. Despite evidence for domain-general variance, our results also suggest domain-specific effects (e.g., Sheinbugs being consistently less correlated with familiar categories then Greebles in Study 3; cars being the most distinct from other object categories while leaves show the largest amount of shared variance with other categories in Study 2). Importantly, the use of multiple categories discourages category-specific or pair-specific explanations of these effects; one could argue that larger correlations between faces and Greebles (versus Ziggerins and Sheinbugs) reflect the fact that the Greebles are more "face-like" in appearance than Sheinbugs (although here we used asymmetrical Greebles that are less face-like), but it is more difficult to argue that Greebles look like cars (see Table 10). In addition, such variability for novel objects suggests that variability in experience and semantics, which account for some of the relations in performance for visual processing with familiar categories (McGugin, Richler, et al., 2012; Van Gulick et al., 2015), are not the only sources of domain-specificity. It may be interesting to investigate at the level of individual differences the role of basic visual properties (like symmetry, curvature or rectangularity) that, on average, account for differences in the recruitment of visual areas (Sasaki, Vanduffel, Knutsen, Tyler, & Tootell, 2005; Nasr et al., 2014; Yue et al., 2014).

We also tested whether general intelligence contributed to performance with faces and cars. Shakeshaft and Plomin (2015) found that performance with both faces and cars shared approximately 2–3% variance with fluid intelligence, and concluded that although one interpretation for this result is that general object recognition is genuinely dissociable from *g* to the same extent as faces, "there is no reason in the literature to suspect that general object recognition is special in this way" (p. 12891). Our results contradict this assumption and add to a growing body of evidence that while performance on any given visual test can be influenced by intelligence, visual object recognition ability itself is distinct from intelligence. Again, we found that although fluid intelligence made a small contribution to NOMT performance (5–8% shared variance), it did not account for shared variance *between* novel object categories. Our re-analysis of Van Gulick et al.'s (2015) Study 2 also found a small contribution of fluid intelligence (4% shared variance) to performance with familiar object categories, but it did not account for the shared variance *between* familiar object categories.

Our results are also consistent with previous work showing that intelligence is often not correlated with performance with faces (our Study 1; Davis et al., 2011) or cars (Van Gulick et al., 2015), and when such correlations are observed they are smaller than those observed for other categories (e.g., 2–3% shared variance with faces in Shakeshaft & Plomin, 2015 and Van Gulick et al., 2015; and with cars in our Study 2 and Shakeshaft & Plomin, 2015). While it remains unclear what factors drive correlations between object recognition and intelligence and lead to smaller or less reliable correlations for some categories (like cars and faces), our results clearly suggest that strong dissociation from IQ is a general feature of object recognition in general, rather than one that is restricted to faces, or even to familiar categories.

The fact that performance dissociates between cars and NOMTs to the same degree as it does between faces and NOMTs highlights the importance of not assuming that performance for all non-face categories reflects category-general object processing mechanisms to a high, and similar, degree. Variability in experience with cars did not account for this dissociation. This reminds us to be cautious in interpreting any given pattern of correlations in terms of evidence for unique abilities. For example, one theoretical position expects face recognition to dissociate to an unusual degree from other objects (e.g., Dennett et al., 2012; Shakeshaft & Plomin, 2015). Another predicts that recognition of different categories should depend on the same domain-general ability in those with high levels of experience and, ideally, little variance in such experience (e.g., Gauthier et al., 2014). Neither of these positions is supported by the present results.

Our work has several limitations. One is that while we used three different novel categories, they were all computer generated and may not represent the full range of possible novel objects that general object recognition could possibly apply to. In that regard, it is reassuring that Study 2 provided converging evidence for domain-general effects that were independent from IQ. In addition, while these are novel categories, similarity to familiar categories for which subjects may have variable experience could have played a role (although this would only reduce our estimates of domain-general variance). A second limitation is that, like prior work (e.g., Shakeshaft & Plomin, 2015; Van Gulick et al., 2015), we probed IQ using a few tasks that are known to load highly on IQ, but do not include a full battery of cognitive tasks to investigate separable IQ-related cognitive abilities. One study found a small but significant correlation between the CFMT and a test of memory for word pairs ($r = 0.17$, Wilmer et al., 2010), and our present study found a significant correlation between the CFMT and Vocabulary ($r = 0.13$). It would be interesting to test whether verbal memory is related to the common variance between NOMTs, especially because of evidence suggesting a continuous hierarchy of functions between ventral visual areas and the medial temporal lobe (Cowell, Bussey, & Saksida, 2010; Shohamy & Turk-Browne, 2013). In contrast, it might be less surprising if performance on subtests that involve individuating objects, such as facial memory subtests present in several intelligence batteries, tapped into the same ability as the NOMTs. Therefore, we do not claim that the ability measured by NOMTs does not contribute to any subtest present in existing batteries of cognitive skill, but that prior work has been generally concerned with the variance that is common across these tasks (g), while we are here concerned with another part of the variance that is common across categories and distinct from g. A third limitation of the present work is that an individual differences based approach does not, by its nature, test detailed accounts of the process by which typical participants learn to recognize novel objects. Such accounts are better tested via experimental studies. For example, prior work with categories of novel objects like the ones used here have suggested that learning is based on image-based representations that come to support generalization to new members of the category from novel viewpoints by varying the threshold of pooled activation in clusters of viewpoint-specific representations of visually similar objects (Edelman, 1995; Tarr & Gauthier, 1998). Other work has found that as we learn to associate non-visual attributes with novel objects, part of the process that takes places during the learning of categories we eventually deem "familiar", left hemisphere performance becomes less viewpoint-dependent, consistent with the idea that the recognition of familiar objects recruits more then perceptual processes (Collins & Olson, 2014; Curby, Hayward, & Gauthier, 2004).

A key result from the present work is the demonstration that it is possible to robustly isolate category-general variance using both NOMTs and FOMTs. Critically, in both cases, this category-general variance dissociated from IQ measures to the point of essential independence. Together, these results provide strong evidence for the existence of mechanisms that contribute to object recognition in a way that is not only domain-general, but, critically, also specific relative to IQ. Beyond these particular results and their implications, the NOMTs that we developed, validated, and normed here are available for use in further research (basic norms are available as supplemental information and via osf.io/6c4m7; more extensive normative data sets, including question-by-question accuracy and reaction time data for all tests described in this paper, are published with Wilmer et al., submitted for publication, and available via osf.io/qygs4). These NOMTs enable the efficient measurement of at least some aspects of domain-object recognition ability, minus familiarity confounds. As such, they provide a potentially valuable contrast for faces and other familiar object categories when testing possible architectures of abilities that support high-level vision.

Acknowledgements

This work was funded by the Temporal Dynamics of Learning Center (National Science Foundation Grant SBE-0542013) and a grant from the National Science Foundation (BCS-1534866).

Appendix A. Recognition questionnaire

Questions 1, 5, 7–9, 11, 12, 14, and 17–20 were used to measures self-report for face recognition ability. Questions 2–4, 6, 10, 13, 15, 16 were used to measures self-report of ability for object recognition *in Study 1. In Study 2, only questions 2–4 were used as a measure of self-report for general visual ability.*

1. On a scale of 1–10 (with 1 being very poor and 10 being very good), where would you place yourself in terms of recognizing faces? (1 = very poor, 10 = very good)
2. How strong is your interest in classifying objects in their various subcategories (e.g., learning about different kinds of insects, plants, shoes, tools, etc.)? (1 = not strong at all, 9 = very strong)
3. How easily do you learn to recognize object visually? (1 = much less than average, 9 = much more than average)
4. Relative to the average person, how much of your typical day involves recognizing things visually? (1 = much less than average, 9 = much more than average)
5. Compared to my peers, I think my face recognition skills are . . . (1 = far below average, 5 = far above average)
6. I can recognize my own baggage at the airport. (1 = never or almost never, 5 = always or almost always)
7. I find it hard to keep track of characters in TV shows or movies. (1 = never or almost never, 5 = always or almost always)
8. When trying to find an acquaintance, I have trouble if they are in a room full of people. (1 = never or almost never, 5 = always or almost always)
9. I find it hard to recognize someone I just met. (1 = never or almost never, 5 = always or almost always)
10. I have trouble recognizing houses I have visited. (1 = never or almost never, 5 = always or almost always)
11. I can recognize well known actors/actresses when watching a movie. (1 = never or almost never, 5 = always or almost always)
12. I notice similarities in the faces of people from the same family. (1 = never or almost never, 5 = always or almost always)
13. I find streets I have often travelled unfamiliar. (1 = never or almost never, 5 = always or almost always)
14. I can recognize famous celebrities in photos or on TV. (1 = never or almost never, 5 = always or almost always)
15. I can recognize particular cats and dogs. (1 = never or almost never, 5 = always or almost always)
16. I'm more likely to identify a car from its license plate number than its overall appearance. (1 = never or almost never, 5 = always or almost always)
17. When I meet someone I pretend to recognize them until their identity is revealed. (1 = never or almost never, 5 = always or almost always)
18. I can recognize casual acquaintances out of context. (1 = never or almost never, 5 = always or almost always)
19. I have trouble recognizing people when they are in uniform. (1 = never or almost never, 5 = always or almost always)
20. I remember the names of people I have met only once or twice. (1 = never or almost never, 5 = always or almost always)

Appendix B. Car recognition questionnaire

1. Please rate yourself on your expertise with cars, considering your interest in, years of exposure to, knowledge of, and familiarity with cars. (1 = very little, 9 = a lot)
2. How important is the domain of cars to you, relative to all other things you are interested in? (1 = very little, 9 = a lot)
3. If you saw a specific car in a TV show, how sure are you that you could recognize that car among similar cars if you were tested the next day? (1 = very little, 9 = a lot)
4. If you were asked to write an essay about different kinds of cars, how extensive and detailed do you think your essay would be? (1 = very little, 9 = a lot).
5. How often do you look at IMAGES of cars in movies, television, or other kinds of documents (books, magazines, online)? (1 = rarely, 9 = often)
6. How often do you read TEXT (in books, magazines, online) that contains information about cars? (1 = rarely, 9 = often)
7. If you are interested in cars, when did this interest begin? (1 = no interest, 6 = 6 or more years ago).

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2017.05.019.

References

Balas, B., & Saville, A. (2015). N170 face specificity and face memory depend on hometown size. *Neuropsychologia, 69*, 211–217.

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

Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence, 24*(1), 25–52.

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cerebral Cortex, 12*(5), 545–551.

Cohen, M. A., Konkle, T., Rhee, J. Y., Nakayama, K., & Alvarez, G. A. (2014). Processing multiple visual objects is limited by overlap in neural channels. *Proceedings of the National Academy of Sciences, 111*(24), 8955–8960.

Cohen, M. A., Nakayama, K., Konkle, T., Stantić, M., & Alvarez, G. A. (2015). Visual awareness is limited by the representational architecture of the visual system. *Journal of Cognitive Neuroscience*.

Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review, 21*(4), 843–860.

Condon, D. M., & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence, 43*, 52–64.

Cor, M. K., Haertel, E., Krosnick, J. A., & Malhotra, N. (2012). Improving ability measurement in surveys by following the principles of IRT: The Wordsum vocabulary test in the General Social Survey. *Social Science Research, 41*(5), 1003–1016.

Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2010). Components of recognition memory: Dissociable cognitive processes or just differences in representational complexity? *Hippocampus, 20*(11), 1245–1262.

Curby, K. M., Hayward, W. G., & Gauthier, I. (2004). Laterality effects in the recognition of depth-rotated novel objects. *Cognitive, Affective, & Behavioral Neuroscience, 4*(1), 100–111.

Davis, J. M., McKone, E., Dennett, H., O'Connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual differences in the ability to recognize facial identity are associated with social anxiety. *PLoS ONE, 6*, e28800.

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). 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.

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.

Edelman, S. (1995). Class similarity and viewpoint invariance in the recognition of 3D objects. *Biological Cybernetics, 72*(3), 207–220.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Services.

Farah, M. J. (1990). *Visual agnosia*. MIT Press.

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*, 164–169.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science, 15*(6), 373–378.

Gauthier, I., McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Van Gulick, A. E. (2014). Experience moderates overlap between object and face recognition, suggesting a common ability. *Journal of Vision, 14*(8), 7.

Hildebrandt, A., Wilhelm, O., Herzmann, G., & Sommer, W. (2013). Face and object cognition across adult age. *Psychology and Aging, 28*, 243–248.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3*, 759–763.

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences, 107*(25), 11163–11170.

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron, 74*(6), 1114–1124.

Li, J., Tian, M., Fang, H., Xu, M., Li, H., & Liu, J. (2010). Extraversion predicts individual differences in face recognition. *Communicative & Integrative Biology, 3*(4), 295–298.

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology, 95*(4), 833.

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*, 17063–17068.

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.

Nasr, S., Echavarria, C. E., & Tootell, R. B. (2014). Thinking outside the box: Rectilinear shapes selectively activate scene-selective cortex. *The Journal of Neuroscience, 34*(20), 6721–6735.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. New York, NY: Psychological Corp.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*(1), 83–92.

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

Sasaki, Y., Vanduffel, W., Knutsen, T., Tyler, C., & Tootell, R. (2005). Symmetry activates extrastriate visual cortex in human and nonhuman primates. *Proceedings of the National Academy of Sciences of the United States of America, 102*(8), 3159–3163.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1), 162.

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences, 112*(41), 12887–12892.

Shohamy, D., & Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology: General, 142*(4), 1159.

Smith, T. W., Marsden, P., Hout, M., & Kim, J. (2013). General Social Surveys, 1972–2012. Chicago, IL: National Opinion Research Center; Storrs, CT: The Roper Center for Public Opinion Research, University of Connecticut.

Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition, 67*(1), 73–110.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.

Van Gulick, A. E., McGugin, R. W., & Gauthier, I. (2015). Measuring nonvisual knowledge about object categories: The Semantic Vanderbilt Expertise Test. *Behavior Research Methods* (in press).

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817.

Wechsler, D., & Hsiao-pin, C. (2011). *WASI-II: Wechsler abbreviated scale of intelligence*. Pearson.

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. http://dx.doi.org/10.1037/a0019972.

Wilmer, J. B. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial vision, 21*(6), 561–579.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: the example of face recognition. *Cognitive Neuropsychology, 29*(5–6), 360–392.

Wilmer, J. B., Richler, J. J., Gauthier, I., & Germine, L. (2017). Large normative datasets for general cognitive ability and visual recognition tests from TestMyBrain.org. *Data in Brief* (submitted for publication).

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

Yue, X., Pourladian, I. S., Tootell, R. B., & Ungerleider, L. G. (2014). Curvature-processing network in macaque visual cortex. *Proceedings of the National Academy of Sciences, 111*(33), E3467–E3475.