Is there a common factor for vision?

Céline Cappe

Laboratory of Psychophysics, Brain Mind Institute,
École Polytechnique Fédérale de Lausanne (EPFL),
Lausanne, Switzerland
Centre de Recherche Cerveau et Cognition,
Université de Toulouse, Toulouse, France

Aaron Clarke

Laboratory of Psychophysics, Brain Mind Institute,
École Polytechnique Fédérale de Lausanne (EPFL),
Lausanne, Switzerland

Christine Mohr

Faculté des Sciences Sociales et Politiques,
Institut de Psychologie, Bâtiment Anthropole, Lausanne,
Switzerland

Michael H. Herzog

Laboratory of Psychophysics, Brain Mind Institute,
École Polytechnique Fédérale de Lausanne (EPFL),
Lausanne, Switzerland

In cognition, common factors play a crucial role. For example, different types of intelligence are highly correlated, pointing to a common factor, which is often called *g*. One might expect that a similar common factor would also exist for vision. Surprisingly, no one in the field has addressed this issue. Here, we provide the first evidence that there is no common factor for vision. We tested 40 healthy students' performance in six basic visual paradigms: visual acuity, vernier discrimination, two visual backward masking paradigms, Gabor detection, and bisection discrimination. One might expect that performance levels on these tasks would be highly correlated because some individuals *generally* have better vision than others due to superior optics, better retinal or cortical processing, or enriched visual experience. However, only four out of 15 correlations were significant, two of which were nontrivial. These results cannot be explained by high intraobserver variability or ceiling effects because test–retest reliability was high and the variance in our student population is commensurate with that from other studies with well-sighted populations. Using a variety of tests (e.g., principal components analysis, Bayes theorem, test–retest reliability), we show the robustness of our null results. We suggest that neuroplasticity operates during everyday experience to generate marked individual differences. Our results apply only to the normally sighted population (i.e., restricted range sampling). For the entire population, including those with degenerate vision, we expect different results.

Introduction

There are many types of intelligence, such as verbal, spatial, and analytical intelligence, which are all highly correlated. For example, verbal intelligence shares 64% of its variability with spatial intelligence (Johnson, Nijenhuis, & Bouchard, 2008). These correlations are assumed to reflect a general intelligence factor, often called *g*.

There are people with eagle-eyed vision and others with lower visual acuity. Superior vision might be due to optical, retinal, or cortical factors but also to differences in visual experiences and attention. For example, if the eyes' optics are blurred, all tasks, for which high-spatial frequency information is important, are affected to the extent that performance may be better in all these tasks for some observers than others. Likewise, differences in the photoreceptor mosaic, attention, and so on may lead to overall perceptual differences. Indeed, basic somatosensory tasks are significantly correlated with each other and, interestingly, even with hearing acuity, likely due to common genetic factors (Frenzel et al., 2012). Hence, we might
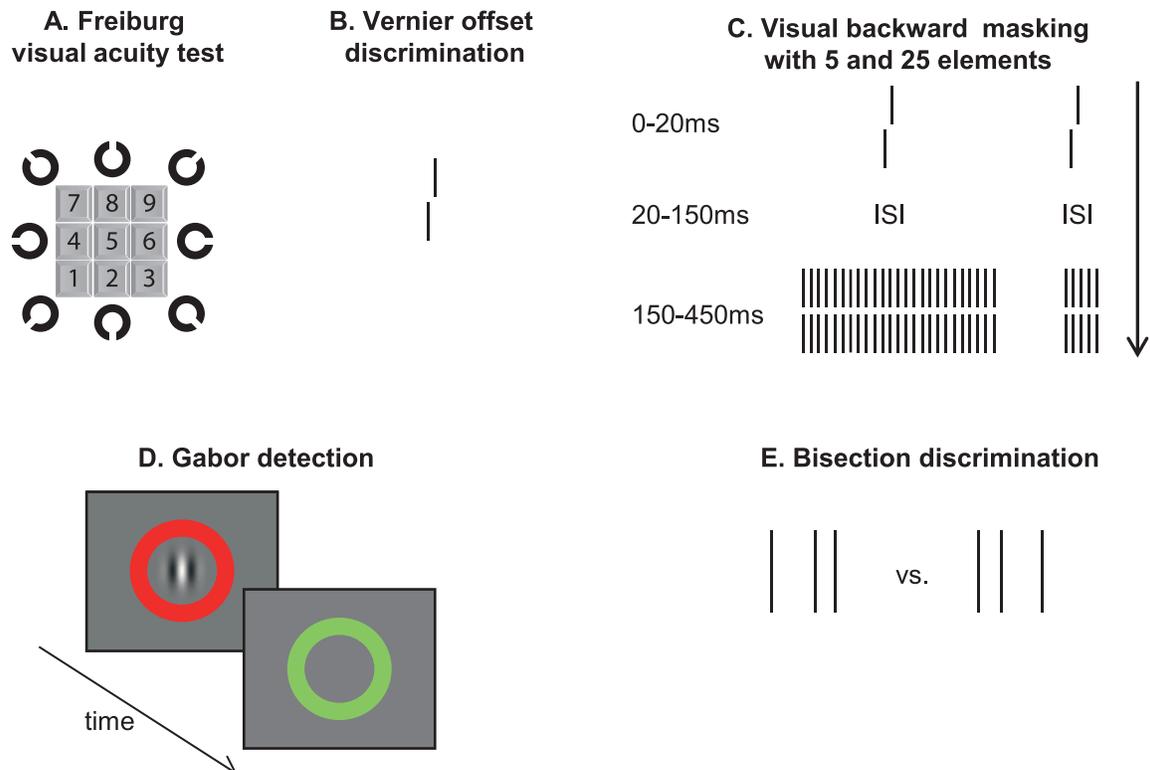
Figure 1. The six basic vision paradigms. (A) FrACT: Participants indicated the gap position in a Landolt ring (eight positions). (B) Vernier offset discrimination: Participants indicated the lower bar's horizontal offset direction relative to the upper bar (left offset in this example). (C) Visual backward masking: Observers performed vernier offset discrimination, with the vernier only briefly presented and subsequently masked for 300 ms. The masking gratings contained either five (BM5) or 25 (BM25) lines. We determined the ISI between vernier termination and grating onset leading to 75% correct responses. (D) Gabor detection: A vertical Gabor appeared in either the first or second interval, indicated by the red and green rings, respectively. Observers indicated in which interval the Gabor was presented. Gabor contrast thresholds for 75% correct responses were determined. (E) Bisection offset discrimination: Observers indicated whether the central line was offset to the left or right of the interval defined by the two outer lines.

expect that in vision there is also a strong common factor leading to high correlations between visual paradigms. This notion is at the very heart of eye doctors' exams where the Snellen eye chart, the Freiburg Visual Acuity Test (FrACT), and the Early Treatment of Diabetic Retinopathy Study chart are assumed to measure general visual acuity and not, for example, just letter-perception acuity.

On the other hand, experience often leads to very specialized visual skills. Radiologists, for example, can detect tumors that are impossible to detect for novices, and cytologists are experts at searching micrographs filled with potentially cancerous cells (Evans et al., 2011). Tennis players show better speed discrimination than the general population (Overney, Blanke, & Herzog, 2008). Also in the laboratory, perceptual learning is usually very specific (but see Aberg, Tartaglia, & Herzog, 2009; Ahissar & Hochstein, 1997; Wright & Sabin, 2007; Xiao et al., 2008). For example, when training improves offset discrimination for a vertical vernier, there is no transfer to a horizontal vernier (Fahle & Edelman, 1993; Spang, Grimsen,

Herzog, & Fahle, 2010). The same holds true for most other visual stimuli such as motion stimuli (Ball & Sekuler, 1987). Thus, the specificity of perceptual learning points to no common factor.

Here, we asked the very general question of whether healthy and well-sighted observers who are good in *one* basic visual task are also good in *other* tasks—that is, whether there is a common visual factor. Our question is not about the interobserver variance in each task; rather, it pertains to whether or not observers vary *consistently* across tasks, with some observers being superior and others inferior in *all* tasks. We chose six basic visual paradigms (Figure 1) and correlated performance.

Materials and methods

Participants

Participants were undergraduate students from the University of Lausanne (Lausanne, Switzerland; $n = 40$,

eight males) who spoke fluent French, with a mean age of 21.1 years (*SD* = 2.0 years, range = 19–27 years). Participants had normal or corrected-to-normal vision as determined by the Freiburg visual acuity test (Bach, 1996); that is, all participants reached a value of >0.8 for at least one eye. Thirty-seven observers were right handed according to a standardized handedness questionnaire (Oldfield, 1971). All observers were naïve to the study's purpose and participated for course credit or financial compensation. The study was conducted in accordance with the guidelines of the declaration of Helsinki. All participants provided written informed consent prior to participation. As indicated by self-report, none of the participants had a previous history of psychiatric or neurological illness.

Apparatus

Verniers and Gabors were presented on a Philips 201B4 Cathode Ray Tube (CRT) monitor (Koninklijke Philips N.V., Amsterdam, Netherlands) driven by a 10-bit Radeon 9200SE graphics card (AMD, Sunnyvale, CA). The monitor was linearized by applying a gamma correction to each color channel individually. The monitor's refresh rate was 100 Hz and its spatial resolution was 1024 × 768 pixels. A viewing distance of 3 m was used in the experiments. The mean luminance was 45 cd/m^2 as determined with a GretagMacbeth Eye-One display 2 colorimeter (GretagMacbeth, Munich, Germany).

Bisection stimuli appeared at the center of a Tektronix 608 display controlled by a computer via fast 16-bit digital-to-analog converters (1 MHz pixel rate; Tektronix, Beaverton, OR). Line elements comprised dots drawn with a dot pitch of 200 μm at a dot rate of 1 MHz. The dot pitch was selected to make the dots slightly overlap; that is, the dot size (or line width) was of the same magnitude as the dot pitch. Stimuli were refreshed at 200 Hz. Luminance was 80 cd/m^2. The room was dimly illuminated (0.5 lux), and background luminance on the screen was below 1 cd/m^2. Subjects observed the stimuli from a distance of 2 m.

Visual paradigms

The basic visual paradigms were administered in the following order. First we administered the FrACT for each eye in one block of 24 trials and then the vernier offset discrimination task. Next, we tested backward masking with five and 25 elements, in two blocks each, with the order of the four blocks fully randomized. Next, in half of the observers we measured Gabor detection in two blocks followed by bisection thresh-

olds; for the other half of the observers it was the other way around.

Vernier offset discrimination

Verniers were presented at a distance of 3 m in a dimly illuminated room. Screen pixels subtended about 18 arcsec at this distance. Verniers were white (100 cd/m^2) on a black background. Verniers consisted of two vertical bars that were 10 arcmin in length and were offset in the horizontal direction. The vernier offset direction was chosen randomly on each trial. Participants indicated the offset direction of the lower bar compared with the upper bar (binary task). Errors were indicated by an auditory signal. In three blocks of 80 trials, we measured vernier offset discrimination thresholds for the shortest possible vernier duration for each observer. This was 10 ms for 39 people and 20 ms for one person. (We first determined vernier offset discrimination for a duration of 40 ms. All observers met the predefined criterion—that is, performance was below 40 arcsec. Then we tested 20 ms and still everyone did well except for one observer. For further details see Herzog & Fahle, 1997; Herzog, Kopmann, & Brand, 2004). The Parametric Estimation by Sequential Testing (PEST) adaptive staircase procedure was used to determine the vernier offset yielding 75% correct responses (Taylor & Creelman, 1967).

Visual backward masking

Vernier duration was 10 ms, as described above (except for one participant whose vernier duration was 20 ms). We used a vernier offset size of 1.15 arcmin for all observers. The vernier was followed by a variable interstimulus interval (ISI; i.e., a blank screen) and then a grating for 300 ms. The grating consisted of either five or 25 verniers (referred to as BM5 and BM25, respectively; BM = backward masking) with zero horizontal offset that were the same length as the target vernier. The horizontal distance between grating elements was about 3.33 arcmin. We varied the ISI adaptively using a staircase procedure (PEST; Taylor & Creelman, 1967). We found the ISI target that yields a performance level of 75% correct responses. (In the figures, we plot stimulus onset asynchrony (SOA) = vernier duration + ISI, rather than ISI.) The starting value of the SOA was 200 ms. Each condition was tested in two blocks of 80 trials each.

Bisection discrimination task

Bisection stimuli consisted of two vertical lines delineating a horizontal interval that was 20 arcmin wide. This interval was bisected into two parts by a central line. The line lengths were 20 arcmin. On a given
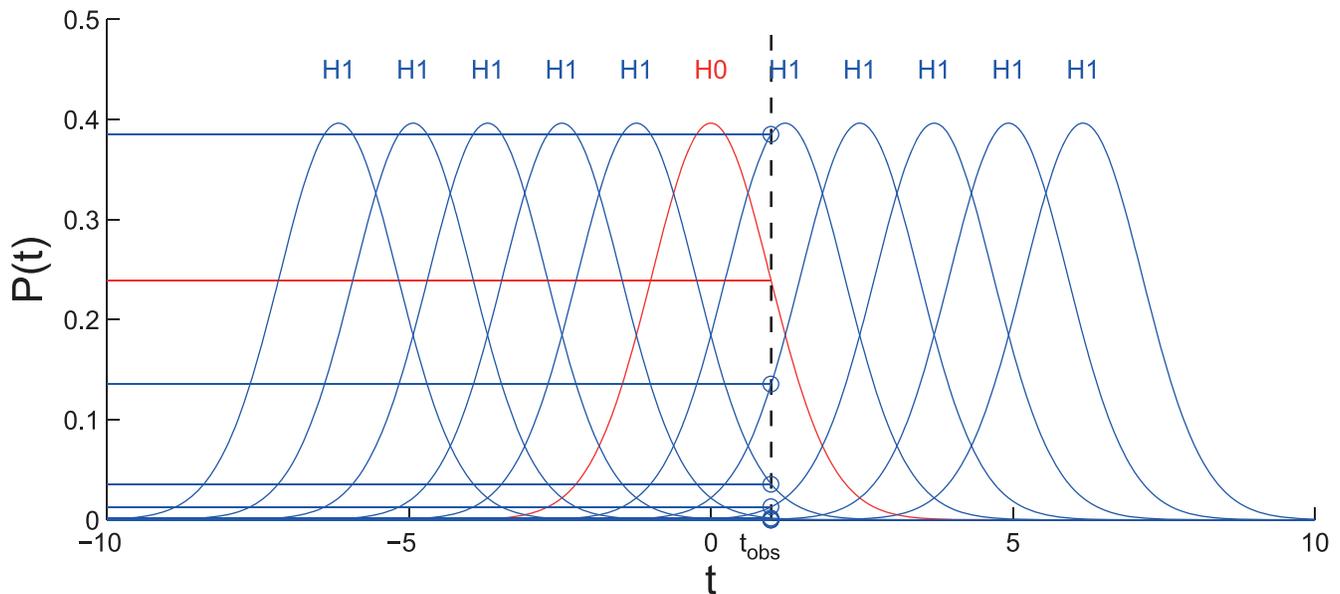
Figure 2. For each variable pair, we computed t_{obs} (vertical, dashed black line) for the hypothesis that the slope equals some arbitrary value (e.g., zero for the null hypothesis). We computed the probability of t_{obs} under the null hypothesis (red curve) and under the alternative hypotheses (blue curves).

trial, the central line was slightly displaced in the direction of one of the two outer lines. The displacement direction was chosen pseudorandomly on each trial. In a binary task, observers indicated the direction of the displacement. We measured the 75% correct bisection thresholds with an adaptive staircase method (PEST; Taylor & Creelman, 1967) and maximum-likelihood estimation of the psychometric function's parameters.

Each trial was initiated with four markers at the corners of the screen presented for 500 ms followed by a blank screen for 200 ms. No fixation point was used in order to prevent observers from judging the position of the central element relative to this fixation point's remembered location. Next, the bisection stimulus was presented for 150 ms at the screen's center. After stimulus presentation, a blank screen appeared for a maximum duration of 3000 ms, during which observers were required to make a response by pressing one of two buttons indicating the bisector's offset direction (right or left). Incorrect responses were followed by an auditory error signal produced by the computer. A new trial was initiated 500 ms after the observer gave a response. Thresholds were determined in two blocks of 80 trials each.

Gabor contrast detection task

We used a two-interval forced-choice contrast detection task. We measured the Gabor contrast detection thresholds at which observers reached 75% correct (two blocks with over 30 trials each) with an adaptive staircase method (PEST; Taylor & Creelman, 1967) and maximum-likelihood estimation of the psychometric function parameters.

The Gabor had a spatial frequency of 4 cycles/deg and appeared either in the first interval within a red ring or in the second interval within a green ring. Observers indicated in which interval the Gabor was presented.

Correlation analysis

We measured the Pearson correlation (r^2) between the mean test values for each participant for each test pair. Kolmogorov–Smirnov tests indicated that all behavioral measures and questionnaire scores were normally distributed. This analysis yielded p-values for each comparison, which we compared against a criterion α-level of 0.05 to determine statistical significance.

Since the traditional frequentist approach of null hypothesis significance testing can only either reject or fail to reject the null hypothesis, we turned to a Bayesian analysis following the approach of Gallistel (2009) in order to determine the likelihood that the null hypothesis was true. Here, we directly compared the probabilities of the null and alternative hypotheses given the data. In order to make this comparison, we needed two quantities: $P(H0|data)$ and $P(H1|data)$. These are (1) the probability of the null hypothesis ($H0$) given the data and (2) the probability of the alternative hypotheses ($H1$) given the data, respectively. Since we were dealing with regression coefficients, we could compute t-statistics for each hypothesis ($t = \sqrt{(df \cdot r^2)/(1 - r^2)}$). For the null hypothesis that

the slope of the regression line is zero, the corresponding *t*-statistic is centered on zero with *n*-2 degrees of freedom (Figure 2, red curve). For the alternative hypotheses, the *t*-statistics may be centered on any value from negative infinity to positive infinity, again with *n*-2 degrees of freedom (Figure 2, blue curves). For the case of the null hypothesis, the posterior probability is computed as

$$P(H0|t_{obs}) = \frac{P(t_{obs}|H0)P(H0)}{P(t_{obs})}.$$

For the alternative hypothesis, the posterior probability is computed as

$$P(H1|t_{obs}) = \frac{\int\limits_{\tau \neq 0} P(t_{obs}|H1 : \mu = \tau)P(H1 : \mu = \tau)d\tau}{P(t_{obs})}.$$

Setting the priors on the two hypotheses to be equal [i.e., $P(H0) = P(H1: \mu = \tau)$] and taking the ratio of these two quantities yields

$$\frac{P(H0|t_{obs})}{P(H1|t_{obs})} = \frac{P(t_{obs}|H0)}{\int\limits_{\tau \neq 0} P(t_{obs}|H1 : \mu = \tau)d\tau}.$$

This is commonly known as the likelihood ratio. The quantity $P(t_{obs}|H0)$ on the right side may be read directly off of the graph in Figure 2, from the red horizontal line. It is given by the value of the *t*-probability density function at the observed *t*-value (t_{obs}). Similarly, the values of $P(t_{obs}|H1: \mu = \tau)$ may also be read off the graph in Figure 2, from the blue horizontal lines. These latter values are continuously distributed under the noncentral *t*-distribution centered on t_{obs}. The integral may be solved numerically to arbitrary precision by approximating it with a Riemann sum. If the posterior ratio is greater than one it implies that the null hypothesis is more probable given the data, and if it is less than one it implies that the alternative hypothesis is more probable given the data. We calculate this quantity for each of our variable pairs and present the results in Figure 3.

Results

We measured performance on visual acuity (FrACT), vernier and bisection offset discrimination, visual backward masking (with five and 25 masking elements), and Gabor contrast detection (Figure 1, Table 1). We found significant Pearson correlations in only four out of the 15 possible task pairs, namely between Gabor contrast detection and visual acuity, between visual backward masking with five and 25 grating elements,

between vernier offset and visual backward masking with five elements, and between vernier offset and bisection discrimination (Figure 3). All other correlations were nonsignificant (Figure 3). The significant correlations between vernier offset discrimination and backward masking and between the two backward masking tasks were expected because these tasks use the same vernier target. Thus, only the correlations between vernier offset and bisection discrimination and between Gabor detection and visual acuity are interesting.

Power

Our very high number of null results is not due to a lack of power. With a sample size of 40 participants, we have sufficient power to detect effects as small as $r^2 = 0.07$, which, according to Cohen (1988, 1992), constitutes a medium effect size. In addition, we did not adjust for multiple comparisons in order to be conservative because we were testing for null effects.

Using a Bayesian approach, we made inferences beyond usual hypothesis testing following the procedures of Gallistel (2009). For each comparison, we compared the probability that the null hypothesis was true (i.e., no relationship between the variables) in relation to the probability that the alternative hypothesis was true (see Correlation analysis for details). This analysis showed that the null hypothesis was more likely than the alternative hypothesis for all comparisons except for the significant cases mentioned above (shown by the likelihood ratios in Figure 3).

Principal components analysis

Because complex, multivariate relationships cannot be detected with pair-wise correlations, we conducted a principal components analysis (PCA). A PCA combines highly correlated variables into one factor. Here, the highest PCA factor explained only 34% of the total variance (Figure 4A). For data sets that *do* have a common factor, the main one or two factors typically explain more than 80% of the variance. For our data set, the main two factors combined explained only 58.6% of the variance; even when including the third factor they explained only 77.2% of the variance (see Figure 4A). Taken together, these findings indicate that there is not a single common factor that can account for the variations in our data.

Ranks

If subjects who performed well in one condition also performed well in other conditions, we would expect
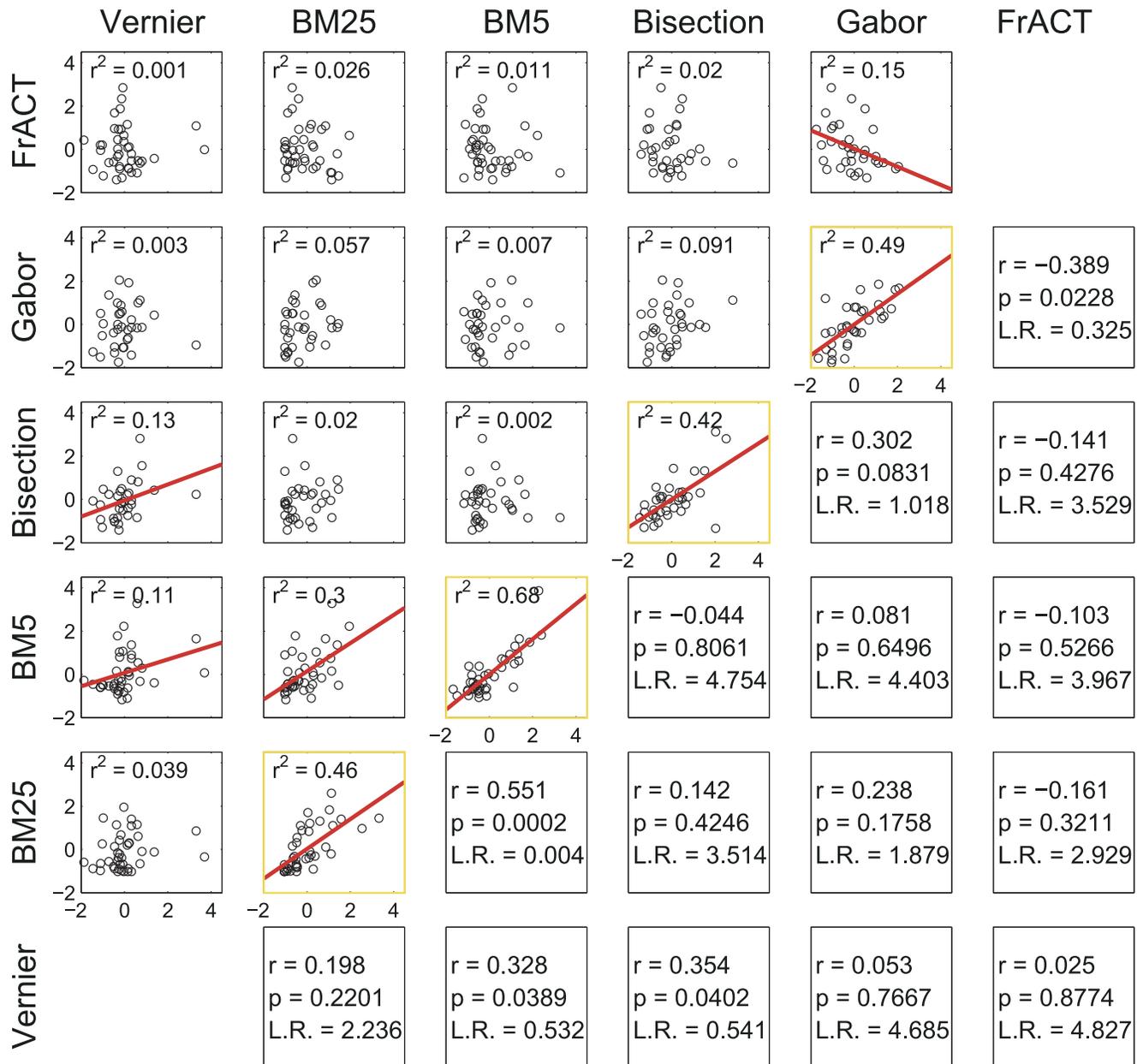
Figure 3. Scatter plots and Pearson correlation values for each pair of tests. Regression lines (red) are plotted only for those pairs that were significantly correlated ($p < 0.05$). To compare values across the various paradigms, we normalized the data by taking *z*-scores (i.e., we subtracted the mean from each value and divided this difference by the standard deviation along each dimension). The variable *r* refers to the Pearson correlation, and *p* is the probability of the null hypothesis that the slope of the regression line is zero. LR denotes the likelihood ratio of the null to the alternative hypothesis (values greater than one imply support for the null hypothesis). Diagonal entries (outlined in orange) show test–retest correlations (i.e., correlations between blocks one and two of the same test). Correlations are high, indicating good test–retest reliability.

their ranks to be consistent from one test to another. In the extreme case, we would expect the best observer to be ranked first in all tests, the second best to be ranked second in all tests, and so on. On the other hand, if performing well in one test is unrelated to performance in another test, then we would expect each observer's ranks over tests to roughly average out to the middle rank (i.e., 20). More than this, we would also expect in

the unrelated case that each observer's ranks should be indistinguishable from a simulated observer with randomly assigned ranks. Figure 5 shows that the observers' ranks differ only slightly from a simulated random observer and that the average rank over observers is very close to 20. This indicates that while the ranks are not completely random (we do find some

	Mean	SD	Range (minimum–maximum)
FrACT	1.35	0.32	0.90–2.25
Gabor	2.75	1.49	0.69–5.38
Bisection	64.35	34.49	22.10–154.45
BM5	68.25	27.72	32.89–159.49
BM25	27.98	16.15	10.00–68.53
Vernier	21.96	12.38	10.10–55.40

Table 1. Mean, standard deviation (SD), and performance range for the six basic visual tasks.
Performance for the six basic visual tasks

correlations between the tasks) they are very close to random.

Test–retest reliability

Our null results are not due to poor test–retest reliability within observers. For the two visual backward masking tasks, the Gabor detection, and the bisection discrimination task, performance was measured twice for each observer in two blocks of 80 trials (Figure 3 shows the average for both blocks). When we calculated Pearson correlations between performance

levels in the two blocks, we found high and significant correlations for each paradigm (see Table 2). There was little intraindividual variability.

In addition to the good test–retest reliability of BM25 and BM5 individually, performance levels for these two tests correlated significantly with each other ($r = 0.551$). This result is not due to the similarity between the masking tasks because the subjective experiences and the corresponding absolute performance levels are very different (Herzog & Koch, 2001). The BM5 mean for the current data is 68 ms ($SD = 27.8$) and the BM25 mean is 28 ms ($SD = 16.2$; Table 1).

Discussion

The main result of our study is the set of amazingly low correlations between performance levels in the various basic visual paradigms. Out of 15 possible correlations, only four were significantly correlated. Even for the four significant results, the correlation coefficients r^2 were very low, in the range 0.1 to 0.3. The Pearson correlation quantifies how much variability in one paradigm is explained by variation in another paradigm. For example, the significant r^2 of 0.11 in
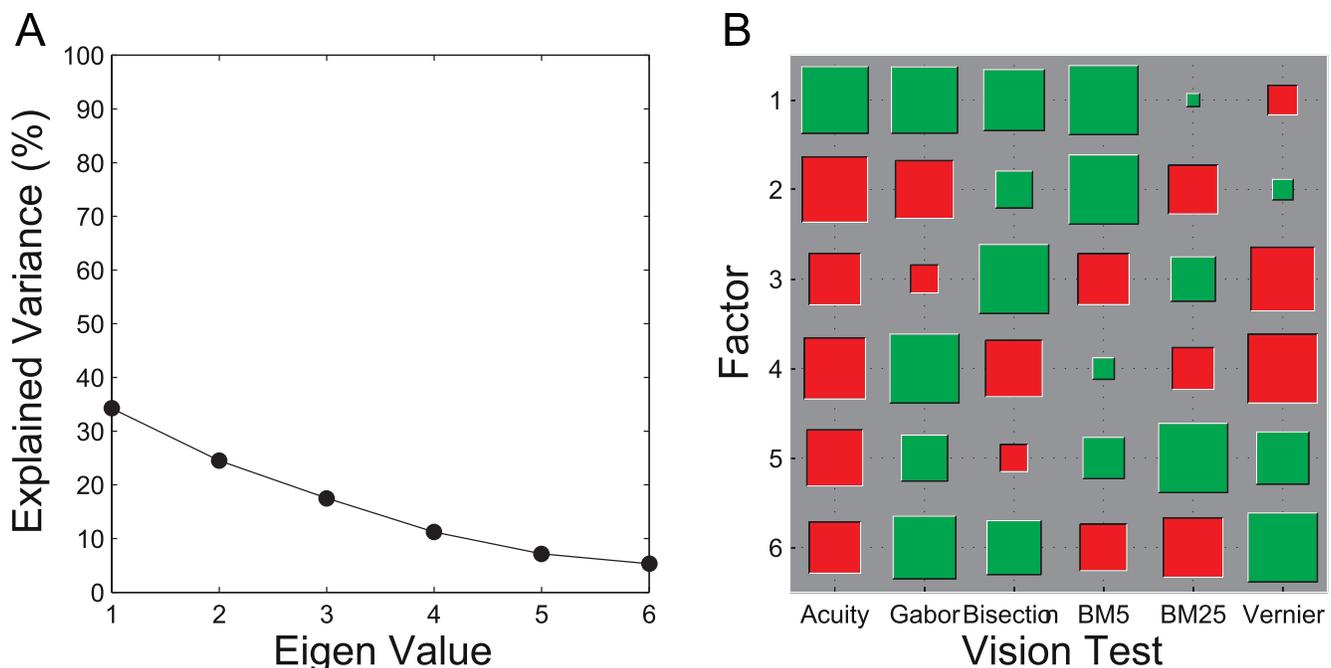

Figure 4. PCA. (A) A PCA combines correlated variables into a single common factor. A factor that is behind many variables has a high eigenvalue. For example, if there is a single factor that explains the variance of all variables, its eigenvalue is 100%. At the other extreme, if there are no common factors, all eigenvalues have the same low value. In our case, three of the eigenvalues are around this lowest value (16.7%) and the remaining eigenvalues are only slightly higher. A common factor usually has an eigenvalue of 80% or more. Clearly, there is no indication for a common factor behind our six tests. (B) Hinton plot showing the loadings for each visual test on each factor. Green squares denote positive loadings while red squares denote negative loadings. Square size indicates loading magnitude.
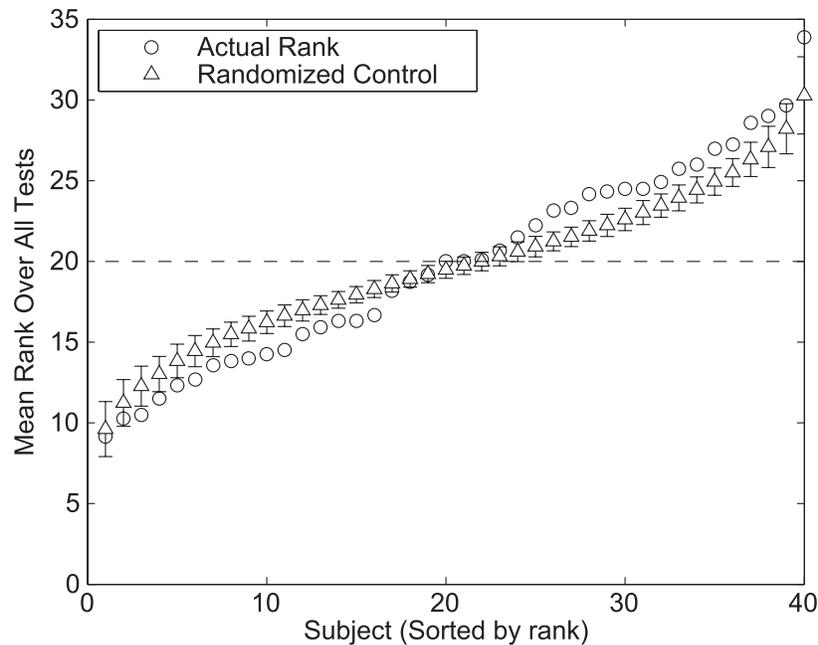
Figure 5. Rank analysis. Sorted mean ranks for each subject (circles) are plotted along with the expected mean ranks assuming complete independence of the ranks on each test (triangles). Error bars denote ±1 *SE*. The dashed line is the expected group average assuming independence of the ranks. Actual ranks are almost random, but with small deviations resulting from the weak correlations we observed between tasks.

Figure 3 indicates that only 11% of the variability in vernier offset discrimination is explained by variability in backward masking with five elements, even though both paradigms required observers to discriminate the very same vernier offsets. As an extreme case of a nonsignificant result, performance in the bisection task has only 0.2% in common with performance in the BM5 backward masking paradigm. Altogether, correlations are low between all the tasks even though all paradigms (except the Gabor task) employ a spatial component.

We expected and found significant correlations between performance levels in the vernier task and the BM5 task as well as between the BM5 and BM25 tasks, likely due to each of these tasks using the vernier target as the task-relevant stimulus. (There was also a trend toward a positive correlation between vernier offset discrimination and BM25.) Besides these "trivial" correlations, only two out of the four remaining significant correlations are "interesting": the correlation between Gabor detection and visual acuity ($r^2 = 0.15$) and the correlation between bisection and vernier discrimination ($r^2 = 0.13$). Both of these correlations are relatively low compared with, for example, the test–retest correlations, which were all higher than $r^2 = 0.42$.

First, as mentioned, our null results cannot be explained by a *lack of power* because we can detect significant differences up to $r^2 = 0.07$, which according to Cohen (1988) is a medium effect size. Moreover, we would have a priori expected the tests to be highly

correlated with r^2 much higher than 0.07. Second, a Bayesian analysis showed that the null hypothesis was more likely than the alternative hypothesis for each of the null findings. Third, our null results cannot be explained by high intraobserver variance because we found high and significant correlations in test–retest conditions in the range of $r^2 = 0.42$ to 0.68 (Table 2). In addition, BM5 and BM25 are strongly correlated with an r^2 of 0.304. Fourth, a PCA showed that there is no common *multivariate* factor behind the tasks. Fifth, null correlations can occur when unmeasured confounding variables correlate positively with one measured variable and negatively with another. It is in general impossible to rule out that there is a hidden cause behind variables. In our case, however, we would have expected the various vision tests to be directly correlated because of their basic nature. Sixth, zero correlations may occur when data are not linearly related. However, inspection of our data shows that with few exceptions (e.g., FrACT vs. vernier) the test pairs were bivariate-normally distributed and did not

	BM25	BM5	Gabor	Bisection
r	0.680284	0.824576	0.699544	0.645619
r²	0.46279	0.67993	0.48936	0.41682
p	0.000001	<0.0000001	0.000002	0.000021

Table 2. Pearson correlations between the two repeated measures for the four visual tests.
Test-re-test correlations

show any salient nonlinearities. Kolmogorov–Smirnov tests on the univariate distributions confirmed their normalcy (vernier: $p = 0.07$; BM5: $p = 0.25$; BM25: $p = 0.20$; bisection: $p = 0.28$; Gabor: $p = 0.79$; FrACT: $p = 0.79$). In addition, our results cannot be explained by outliers. Seventh, we have sufficient variance in our student data to avoid ceiling effects. The variance in our student sample is similar to other well-sighted populations. In the visual acuity test, the mean was 1.35 and *SD* was 0.32 (Table 1), which is comparable with a much larger sample of 817 student observers from our laboratory database (mean = 1.31, *SD* = 0.34) and with 138 healthy participants from the general population who served as age- and education-matched controls in experiments researching schizophrenia (mean = 1.37, *SD* = 0.36). Hence, our student sample can be considered to be representative of the population of normal, well-sighted or corrected-to-normal observers.

It is surprising that visual acuity (FrACT) correlated so poorly with performance in the other visual paradigms. Importantly, however, we do *not* claim that there are no correlations between visual acuity tests and basic visual skills in general. Here, we tested only observers with good and corrected-to-normal vision (i.e., observers with values larger than 0.8 in the FrACT). If we had tested the entire population, including people with low vision, leading to a full range of acuity values, there would have been much stronger correlations. For the entire population the FrACT correlates very well with other classical eye tests such as the Early Treatment of Diabetic Retinopathy Study chart ($r^2 = 0.92$; Kurtenbach, Langrová, Messias, Zrenner, & Jägle, 2013). Thus, the FrACt can be considered to be a "good" visual test. For this reason, we do not doubt the validity of acuity tests. Our null correlations hold true only in the *normal* population with people having average to good visual acuity (restricted range sampling). Similar considerations hold true also for vernier acuity and backward masking. McKee, Levi, and Movshon (2003) investigated visual acuity and vernier acuity in healthy and amblyopic observers. Whereas vernier acuity varied strongly with visual acuity across the entire population, there were no obvious correlations for the 68 well-sighted observers in the study (McKee et al., 2003).

Many factors determine visual perception. Among them are the quality of the eye's optics, the density of the retinal photoreceptor array, the quality of encoding in the primary visual cortex, the ability of higher-level areas to read out signals from early visual areas, attention, cognitive factors, and decision making. It seems likely that superiority in any of these "unspecific" factors leads to superior performance in many tasks, and this may explain why there are eagle eyes and average eyes. In light of this, however, it is surprising that we observed so few significant correlations.

Where do the large differences in these visual tests come from? We suggest the differences may be explained by the individual experiences of observers (assuming there are no innate visual abilities for the paradigms tested here). We propose that everyday perceptual learning leads to very specialized skills that do not transfer to similar paradigms. The situation may be different for more complex stimuli, for which transfer has been found. For example, it is surprising that action video gaming is so much more powerful in generating transfer in basic visual paradigms such as Gabor detection than are the visual experiences of everyday life (Bavelier, Green, Pouget, & Schrater, 2012; Li, Polat, Makous, & Bavelier, 2009; Polat et al., 2012). The transfer may potentially be explained by the complex stimuli and by the prolonged, heightened attention and strong arousal that video gaming brings about.

An alternative explanation is that hyperacuity tasks, such as vernier and bisection acuity, involve mainly cortical processing, while the FrACT and contrast detection involve mainly retinal-level processing. However, there is no consensus about the main processing sites for our visual stimuli and it is very unlikely that one stimulus is processed at *only* one of these two levels. In addition, there are many other factors that contribute to vision, such as the optical apparatus, attention, decision making, and so on. Moreover, bisection acuity and the masked vernier task (BM5) showed the lowest pair-wise correlation coefficient ($r^2 = 0.002$), even though both could be posited as being more related to cortical processing.

Our results stand in contrast to those of other studies were high correlations were found between paradigms that one would naïvely imagine being much more variable than the basic spatial tasks we used. Palmer and Griscom (2012), for example, asked observers to rate their preferences for harmony in color and in musical pieces and found surprisingly high correlations in the range of 0.6. Similarly, ratings on aesthetics between emotion and color were in the range of $r^2 = 0.64$ (Palmer & Schloss, 2010). Recent studies showed that intelligence quotient and performance levels on four different associative memory tasks were all positively and highly correlated (all $r^2 > 0.36$; Ratcliff, Thapar, & McKoon, 2011). Interestingly, in the auditory domain it seems to be that there is a common factor for basic audition (Kidd, Watson, & Gygi, 2007).

Whereas there are research fields that regularly observe high correlations between measures, there are also others that observe very low correlations. In a recent study, the strength of two illusions (Ponzo and Ebbinghaus illusions) correlated with V1 surface area (Schwarzkopf, Song, & Rees, 2011). In this same study, however, the two illusions did not significantly correlate with each other ($r^2 = 0.06$, $n = 30$). In 490 observers,

it was found that many spatial illusions, including the Müller-Lyer and Ebbinghaus illusions, do not share a common spatial factor (Coren & Porac, 1987). Furthermore, people from different cultures seem to perceive the world differently. For example, the strength of the Müller-Lyer illusion is almost zero in the San foragers of the Kalahari but very pronounced in Westerners (Henrich, Heine, & Norenzayan, 2010). In the same line as in the present study, Goodbourn et al. (2012) tested four different magnocellular tasks and showed that these tests are not highly correlated. In addition, they found that among these tasks, only one pair shared more than 4% of variance and that correlations between these tasks and a nonmagnocellular task were similarly low. In combination, these and our results suggest that the visual system is highly modular, with each module specializing in a different form of processing, but with little overlap between modules.

In summary, we showed that there are very few significant correlations among basic visual tests, indicating that there is no obvious general factor underlying vision. We propose that experience shapes vision in a very specific manner. Our results apply only to people with good vision and do not challenge the value of eye tests for the general population. In general, the interesting question arises: Under which conditions do high correlations occur between different tests and when are lower correlations found?

Keywords: vision, detection, discrimination

Acknowledgments

Commercial relationships: none.
Corresponding author: Celine Cappe.
Email: celine.cappe@cerco.ups-tlse.fr.
Address: Centre de Recherche Cerveau et Cognition (CerCo), Toulouse, France.

References

Aberg, K. C., Tartaglia, E. M., & Herzog, M. H. (2009). Perceptual learning with chevrons requires a minimal number of trials, transfers to untrained directions, but does not require sleep. *Vision Research, 49,* 2087–2094.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387,* 401–406.

Bach, M. (1996). The Freiburg visual acuity test—Automatic measurement of visual acuity. *Optometry and Vision Science, 73,* 49–53.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research, 27,* 953–965.

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience, 35,* 391–416.

Cohen, C. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, C. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Coren, S., & Porac, C. (1987). Individual differences in visual-geometric illusions: Predictions from measures of spatial cognitive abilities. *Perception and Psychophysics, 41,* 211–219.

Evans, K. K., Cohen, M. A., Tambouret, R., Horowitz, T., Kreindel, E., & Wolfe, J. M. (2011). Does visual expertise improve visual recognition memory? *Attention, Perception and Psychophysics, 73,* 30–35.

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research, 33,* 397–412.

Frenzel, H., Bohlender, J., Pinsker, K., Wohlleben, B., Tank, J., Lechner, S. G., et al. (2012). A genetic basis for mechanosensory traits in humans. *PLoS Biology, 10*(5), e1001318.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review, 116,* 439–453.

Goodbourn, P. T., Bosten, J. M., Hogg, R. E., Bargary, G., Lawrance-Owen, A. J., & Mollon, J. D. (2012). Do different "magnocellular tasks" probe the same neural substrate? *Proceedings of the Royal Society B: Biological Sciences, 279,* 4263–4271.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences, 33,* 61–83.

Herzog, M., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research, 37,* 2133–2141.

Herzog, M. H., & Koch, C. (2001). Seeing properties of an invisible object: Feature inheritance and shine-through. *Proceedings of the National Academy of Sciences, USA, 98,* 4271–4275.

Herzog, M. H., Kopmann, S., & Brand, A. (2004). Intact figure-ground segmentation in schizophrenia. *Psychiatry Research, 129,* 55–63.

Johnson, W., Nijenhuis, J. T., & Bouchard, T. J., Jr. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence, 36,* 81–95.

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America, 122,* 418–435.

Kurtenbach, A., Langrová, H., Messias, A., Zrenner, E., & Jägle, H. A. (2013). Comparison of the performance of three visual evoked potential-based methods to estimate visual acuity. *Documenta Ophthalmologica, 126,* 45–56.

Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience, 12,* 549–551.

McKee, S. P., Levi, D. M., & Movshon, J. A. (2003). The pattern of visual deficits in amblyopia. *Journal of Vision, 3*(5):5, 380–405, http://www.journalofvision.org/content/3/5/5, doi:10.1167/3.5.5. [PubMed] [Article]

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Overney, L. S., Blanke, O., & Herzog, M. H. (2008). Enhanced temporal but not attentional processing in expert tennis players. *PLoS One, 3,* e2380.

Palmer, S. E., & Griscom, W. S. (2012). Accounting for taste: Individual differences in preference for harmony. *Psychonomic Bulletin and Review, 20,* 453–461.

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences, USA, 107,* 8877–8882.

Polat, U., Schor, C., Tong, J. L., Zomet, A., Lev, M., Yehezkel, O., et al. (2012). Training the brain to overcome the effect of aging on the human eye. *Scientific Reports, 2,* 278.

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology. General, 140,* 464–487.

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience, 14,* 28–30.

Spang, K., Grimsen, C., Herzog, M. H., & Fahle, M. (2010). Orientation specificity of learning vernier discriminations. *Vision Research, 50,* 479–485.

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *The Journal of the Acoustical Society of America, 41,* 782–787.

Wright, B. A., & Sabin, A. T. (2007). Perceptual learning: How much daily training is enough? *Experimental Brain Research, 180,* 727–736.

Xiao, L. Q., Zhang, J. Y., Wang, R., Klein, S. A., Levi, D. M., & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Current Biology, 18,* 1922–1926.