

Individual differences in the perception of visual illusions are stable across eyes, time, and measurement methods

Aline F. Cretenoud

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



Lukasz Grzeczowski

Laboratory of Psychophysics, Brain Mind Institute,
Ecole Polytechnique Fédérale de Lausanne (EPFL),
Lausanne, Switzerland
Active Perception and Cognition, Department of
Psychology, Humboldt-Universität zu Berlin, Berlin,
Germany



Marina Kunchulia

Institute of Cognitive Neurosciences, Free University of
Tbilisi, Tbilisi, Georgia
Laboratory of Vision Physiology, Ivane Beritashvili Center
of Experimental Biomedicine, Tbilisi, Georgia



Michael H. Herzog

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



Vision scientists have tried to classify illusions for more than a century. For example, some studies suggested that there is a unique common factor for all visual illusions. Other studies proposed that there are several subclasses of illusions, such as illusions of linear extent or distortions. We previously observed strong within-illusion correlations but only weak between-illusion correlations, arguing in favor of an even higher multifactorial space with—more or less—each illusion making up its own factor. These mixed results are surprising. Here, we examined to what extent individual differences in the perception of visual illusions are stable across eyes, time, and measurement methods. First, we did not find any significant differences in the magnitudes of the seven illusions tested with monocular or binocular viewing conditions. In addition, illusion magnitudes were not significantly predicted by visual acuity. Second, we observed stable individual differences over time. Last, we compared two illusion measurements, namely an adjustment procedure and a method of constant stimuli, which both led to similar individual differences. Hence, it is unlikely that the individual differences in the perception of visual illusions arise from instability across eyes, time, and measurement methods.

Introduction

Common factors are ubiquitous in everyday life. For example, there seems to be a strong common factor for cognition in healthy aging, that is, cognitive abilities are reliably affected with age (e.g., [Baltes & Lindenberger, 1997](#); [Kiely & Anstey, 2017](#); [Lindenberger & Ghisletta, 2009](#)). Age-related changes of different cognitive functions, such as perceptual speed and reasoning skills, were indeed reported to significantly correlate. In analogy, a strong common factor for vision may be expected, i.e., it may be that an individual who performs better in one visual task compared to other individuals also performs better in other visual tasks, suggesting that there is a single monolithic structure underlying vision. However, the space underlying vision seems to be multifactorial, i.e., there seems to be no unique common factor for vision (for reviews, see [Mollon, Bosten, Peterzell, & Webster, 2017](#); [Tulver, 2019](#)). For example, [Cappe, Clarke, Mohr, & Herzog \(2014\)](#) only observed weak correlations between the performance in six basic visual paradigms, such as visual acuity and contrast detection, suggesting that

Citation: Cretenoud, A. F., Grzeczowski, L., Kunchulia, M., & Herzog, M. H. (2021). Individual differences in the perception of visual illusions are stable across eyes, time, and measurement methods. *Journal of Vision*, 21(5):26, 1–20, <https://doi.org/10.1167/jov.21.5.26>.



an individual with good performance in one task does not necessarily show good performance in other tasks. A principal component analysis revealed a first component explaining only 34% of the variability in the data (but see [Bosten, Goodbourn, Bargary, Verhallen, Lawrance-Owen, Hogg, & Mollon, 2017](#)). Likewise, several factors were suggested to underlie individual differences in hue scaling ([Emery, Volbrecht, Peterzell, & Webster, 2017a](#); [Emery, Volbrecht, Peterzell, & Webster, 2017b](#)), oculomotor tasks ([Bargary, Bosten, Goodbourn, Lawrance-Owen, Hogg, & Mollon, 2017](#)), and binocular rivalry (e.g., [Brascamp, Becker, & Hambrick, 2018](#)).

Common factors were proposed for visual illusions. For example, [Thurstone \(1944\)](#) observed a factor underlying geometric illusions. Similarly, [Roff \(1953\)](#) computed a factor analysis on 70 perceptual measures, which resulted in a single factor associated with visual illusions (see also [Aftanas & Royce, 1969](#)). However, illusions were also shown to be more heterogeneous. For example, [Coren, Girgus, Erlichman, & Hakstian \(1976\)](#) claimed that illusions belong to two classes, namely the illusions of extent and the illusions of shape or direction (see also [Robinson, 1968](#)). Likewise, [Taylor \(1974, 1976\)](#) reported that several illusion measures were best represented by a four-factor model including illusions of length judgments and distortions of parallelism.

We previously observed strong within-illusion correlations but only weak between-illusion correlations, suggesting that there are illusion-specific factors. For example, the susceptibility to the Müller-Lyer illusion was only weakly correlated with the susceptibility to the Ponzo illusion ([Cretenoud, Francis, & Herzog, 2020](#); [Cretenoud, Grzeczowski, Bertamini, & Herzog, 2020](#)). However, several variants of the Ebbinghaus illusion, differing in color, texture, shape, or contrast, were strongly intercorrelated, arguing in favor of a common mechanism for the Ebbinghaus illusion ([Cretenoud, Karimpur, Grzeczowski, Francis, Hamburger, & Herzog, 2019](#)). The same was true for nine other illusions tested with different contrast and orientation conditions.

These mixed results, i.e., whether there is one unique or several specific factors for visual illusions, may come from discrepancies in the data analysis and interpretation of the results or in the experimental design. For example, some studies used an adjustment procedure whereas others used a method of constant stimuli ([Peterzell, 2020](#)). Also, it may be that illusion magnitudes are not stable in space and time, leading to different results across studies.

Here, we wondered how stable individual differences in the perception of illusions are. We first investigated whether illusory percepts are reliable interocularly and whether illusion magnitudes differ as a function of visual acuity. Second, the susceptibility to several illusions was measured at different time points within

a month to evaluate how stable individual differences are over time. Last, we tested two illusions, namely the Ebbinghaus and Müller-Lyer illusions, with both an adjustment procedure and a method of constant stimuli (2-AFC) to determine whether reliable individual differences are observed with different measurement methods.

Experiment 1

Monocular viewing condition was shown to result in a significantly weaker magnitude of an actual, real-world Ponzo illusion compared to binocular viewing condition, because of a reduced perception of depth cues following the elimination of stereopsis ([Leibowitz, Brislin, Perlmutter, & Hennessy, 1969](#)). Here, we first examined whether the susceptibility to several two-dimensional visual illusions is different between eyes and for monocular and binocular presentations. Second, we investigated the role of visual acuity in the perception of visual illusions.

Methods

Participants

Fifteen students and collaborators of the Ecole Polytechnique Fédérale de Lausanne (EPFL, Switzerland) participated in this experiment (five females; mean age, 24 years; age range, 18–52 years). Participants were naïve to the purpose of the experiment. Prior to the experiment, participants signed informed consent. Both monocular and binocular visual acuities were measured in a random order using the Freiburg visual acuity test ([Bach, 1996](#)). Participants were paid 20 Swiss Francs per hour. Procedures were conducted in accordance with the Declaration of Helsinki, except for preregistration (§35), and were approved by the local ethics committee.

Apparatus

A BenQ XL2420T LCD monitor driven by a Windows-PC using Matlab (MathWorks Inc., Natick, MA, USA) and the Psychophysics toolbox ([Brainard, 1997](#); [Pelli, 1997](#)) was used. Stimuli were presented at a 1920 × 1080 pixel resolution with a 60 Hz refresh rate. The distance to the screen was approximately 60 cm. During the monocular testing, participants covered one of their eyes with an eyepatch. Participants adjusted stimuli with a Logitech LS1 computer mouse. The experiment was conducted in the Laboratory of Psychophysics at EPFL, Switzerland.

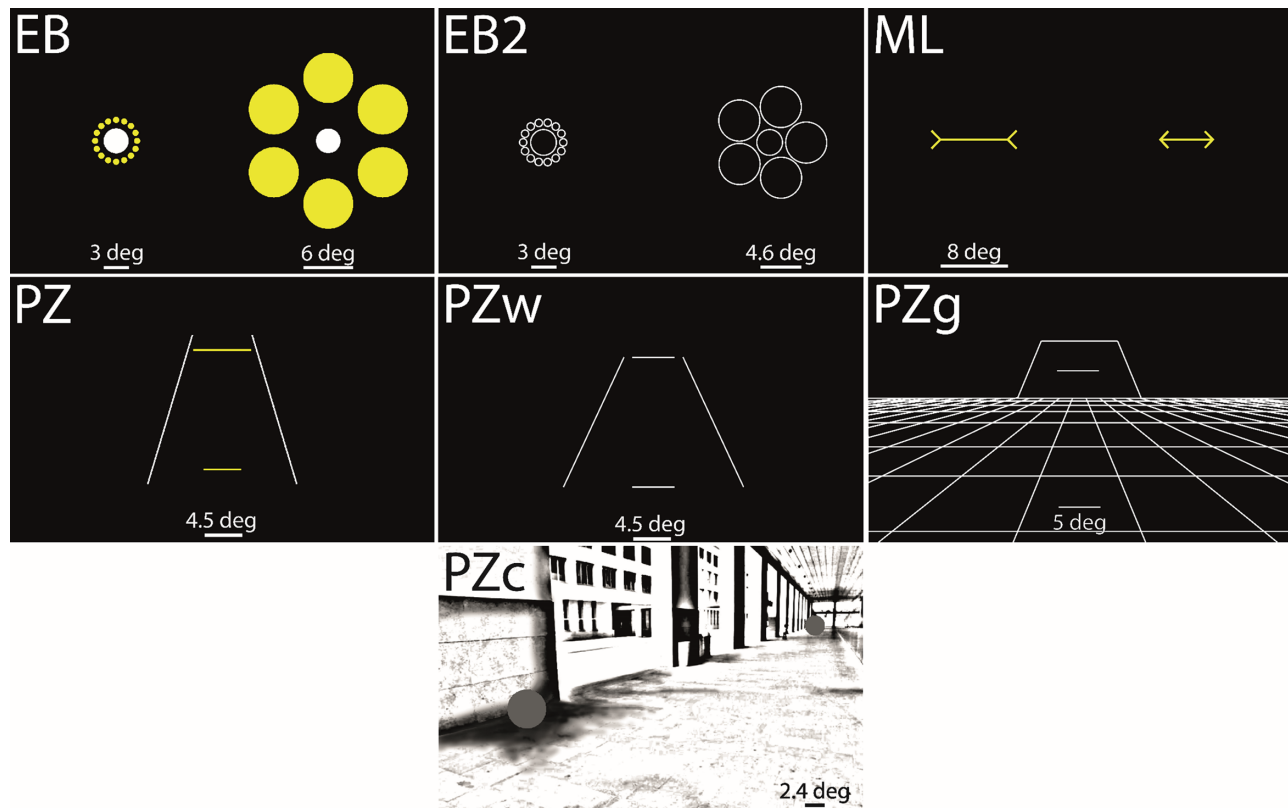


Figure 1. The seven illusions tested in [Experiment 1](#): two variants of the Ebbinghaus (EB and EB2), the Müller-Lyer, and four variants of the Ponzo (PZ, PZw, PZg, and PZc). For each stimulus, participants were asked to adjust the size of a target to match the size of a reference element. For example, in the Ebbinghaus stimuli (EB and EB2), participants adjusted the size of the right central disk to match it with the left central disk using a computer mouse. Each stimulus was tested monocularly (left and right) and binocularly.

Stimuli

Each participant was tested with seven illusions ([Figure 1](#)): two variants of the Ebbinghaus (EB and EB2), a Müller-Lyer (ML), and four variants of the Ponzo (PZ; a wider variant, PZw; a grid variant, PZg; and a corridor variant, PZc) illusion. Note that the corridor variant of the Ponzo illusion (PZc) has also previously been called a Ponzo “hallway” illusion ([Grzeczowski, Clarke, Francis, Mast, & Herzog, 2017](#); [Grzeczowski, Roinishvili, Chkonina, Brand, Mast, Herzog, & Shaqiri, 2018](#)).

Each illusion was tested monocularly (left and right, separately) and binocularly. A method of adjustment was used, that is, participants adjusted the size of a target to match the size of a reference on the screen by moving the computer mouse on the horizontal axis. In the Ebbinghaus illusions (EB and EB2), participants adjusted the size of the right central disk to match the left central disk in size. Participants adjusted the length of the right horizontal segment to match the length of the left horizontal segment in the ML. In three variants of the Ponzo illusion (PZ, PZw, and PZg), participants adjusted the length of the upper horizontal

segment to match the length of the lower horizontal segment. In the corridor variant of the Ponzo illusion (PZc), participants adjusted the size of the disk in the lower-left quadrant to match the size of the disk in the upper-right quadrant. The luminance was approximately 1 cd/m² for the black background, 30 cd/m² for gray, 146 cd/m² for yellow, and 176 cd/m² for white. For a more detailed description of the different illusions, please refer to [Grzeczowski et al. \(2017\)](#).

Procedure

The experimenter first explained the task to the participants who completed one trial of each illusion binocularly to familiarize with the method of adjustment. The 21 conditions (7 illusions × 3 eye conditions) were then tested twice (42 trials in total). The illusions were presented in the same order to all participants, but the order of eye conditions, that is, left, right, and binocular, was randomly set for each illusion and for each participant. The two trials of the same condition were always presented in a sequential manner, that is, one trial after the other. There was no time restriction and no feedback.

Data analysis

Analyses were performed in Matlab (Mathworks Inc.) and R (R Core Team, 2018). We assessed intrarater reliabilities for each illusion, that is, the within-individual variation of illusion magnitudes across trials, by computing intraclass correlations (ICC) among the six adjustments of each illusion (3 eye conditions \times 2 trials). First introduced by Fisher (1992) as an extension of the Pearson correlation coefficient, the concept of ICC was later developed as a measure of reliability within a class of data (Bartko, 1966; Shrout & Fleiss, 1979) rather than between different classes (the correlation between two different classes of data is usually computed as a Pearson correlation). ICCs are based on the analysis of variance (ANOVA) and therefore assume normally distributed data. In short, an ICC is computed as a ratio between the variance of interest (e.g., between-individual variance) and the total variance, i.e., the variance of interest and unwanted variance (e.g., within-individual variance or instrumental variation). The larger the ICC coefficient, the more reliable the data.

Several types of ICCs were developed to fit different experimental situations (e.g., Koo & Li, 2016; Liljequist, Elfving, & Skavberg Roaldsen, 2019). Here, we computed intraclass correlations of type (3,1) or ICC_{3,1}. The first subscript indicates that we computed a two-way mixed effects model (i.e., model 3), in which a random sampling of participants is assumed, whereas biases are assumed to be fixed (i.e., the only measure of interest is the illusion magnitude, which was measured several times to assess the intrarater reliability). The second subscript indicates the type of selection used, i.e., each data point either represents a single measurement (i.e., type 1) or an average across several measurements (i.e., type k). We computed a 95% confidence interval around the correlation coefficient r , as suggested in Shrout & Fleiss (1979).

For each participant and each condition (7 illusions \times 3 eye conditions = 21 conditions), we then averaged the adjusted values from both trials. The size of the reference element in each illusion was subtracted from the averaged values. Hence, the illusion magnitude is expressed as a size difference compared to the reference with positive and negative values indicating over- and under-adjustments, respectively. Correlations were computed between each pair of conditions. Cohen (1988) considered correlation coefficients of 0.1, 0.3, and 0.5 as small, medium, and large effect sizes, respectively. According to a meta-analysis by Gignac and Szodorai (2016), effect sizes of 0.1, 0.2, and 0.3 are, however, considered as small, medium, and large, respectively.

To account for random variations in the baseline of participants and illusions, we computed mixed effects models (lmer R package). The fixed effects were eye

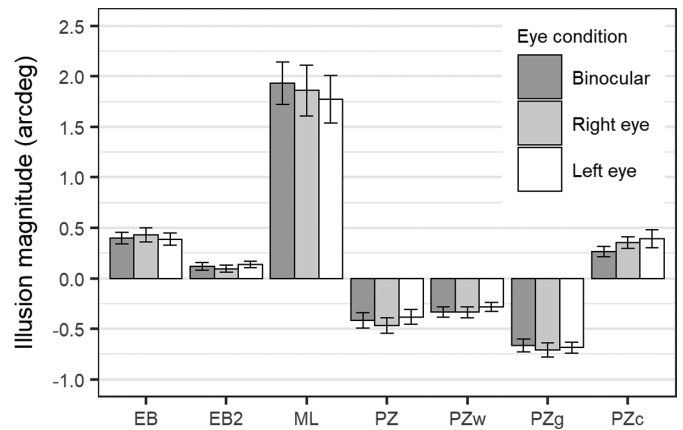


Figure 2. Illusion magnitudes \pm SEM (standard error of the mean) in Experiment 1 as a function of the eye condition (dark grey: binocular; light grey: monocular right; white: monocular left). Positive and negative magnitudes indicate over- and under-adjustments compared to the reference, respectively.

condition, visual acuity, and sex. We accounted for the random effects of participants and illusions (random intercepts). The significance of each predictor in the model was assessed by computing likelihood ratio tests, which express the relative likelihood of the data given two competing models. The effect size was computed as a measure of explained variance with the random effect structure included (MuMIn R package).

Results

Intrarater reliabilities

For each illusion, we assessed intrarater reliability by computing an intraclass correlation. All ICC coefficients were significant even after correcting for inflated family-wise errors (EB: ICC coef. = 0.823, 95% CI [0.710, 0.913], $p < 0.001$; EB2: ICC coef. = 0.705, 95% CI [0.550, 0.846], $p < 0.001$; ML: ICC coef. = 0.465, 95% CI [0.285, 0.680], $p < 0.001$; PZ: ICC coef. = 0.489, 95% CI [0.309, 0.699], $p < 0.001$; PZw: ICC coef. = 0.628, 95% CI [0.458, 0.797], $p < 0.001$; PZg: ICC coef. = 0.686, 95% CI [0.527, 0.834], $p < 0.001$; PZc: ICC coef. = 0.669, 95% CI [0.506, 0.824], $p < 0.001$). The effect sizes are large according to Gignac and Szodorai (2016) and medium to large according to Cohen (1988), which suggests that the adjustments were consistent across eye conditions.

Illusion magnitudes

Mean illusion magnitudes are shown in Figure 2 and summarized in Supplementary Table S1. There were strong between-illusion differences, whereas only weak differences were observed between eye conditions. The

		EB			EB2			ML			PZ			PZw			PZg			PZc			
		Binocular	Right	Left	Binocular	Right	Left	Binocular	Right	Left	Binocular	Right	Left	Binocular	Right	Left	Binocular	Right	Left	Binocular	Right	Left	
EB	Binocular		.94	.94	.57	.35	.56	.20	.69	.50	-.20	-.15	.03	-.01	-.07	-.27	-.16	-.24	-.07	.68	.77	.84	
	Right			.88	.49	.39	.52	.32	.66	.42	-.14	-.19	-.15	.07	-.13	-.41	-.02	-.23	-.08	.57	.66	.75	
	Left				.58	.40	.61	.11	.61	.54	-.15	-.13	-.03	.09	-.12	-.25	-.15	-.29	-.11	.61	.69	.82	
EB2	Binocular					.81	.70	.13	.45	.48	-.41	-.22	-.25	-.11	.06	-.25	-.21	-.31	-.25	.44	.62	.43	
	Right						.76	.01	.10	.37	-.55	-.40	-.63	-.29	-.30	-.64	-.19	-.55	-.44	.10	.23	.07	
	Left							-.36	.04	.36	-.75	-.65	-.32	-.39	-.52	-.66	-.51	-.72	-.44	.44	.50	.40	
ML	Binocular								.59	.30	.45	.29	-.16	.32	.46	.26	.47	.43	.28	-.14	.05	.08	
	Right									.55	.25	.26	.16	.36	.51	.20	.33	.36	.27	.36	.51	.61	
	Left										-.28	-.16	-.04	-.05	-.05	-.03	.09	-.16	-.10	.06	.22	.30	
PZ	Binocular											.80	.31	.74	.70	.61	.59	.83	.60	-.28	-.29	-.09	
	Right												.47	.63	.74	.57	.59	.81	.55	-.27	-.22	-.10	
	Left													.33	.54	.69	.16	.50	.57	.20	.03	.12	
PZw	Binocular														.67	.54	.79	.70	.74	-.22	-.21	.14	
	Right															.77	.58	.86	.72	-.09	-.06	-.02	
	Left																.38	.70	.65	-.11	-.17	-.05	
PZg	Binocular																	.71	.71	-.57	-.49	-.22	
	Right																			.80	-.38	-.28	-.14
	Left																				-.33	-.30	-.05
PZc	Binocular																				.87	.76	
	Right																					.86	
	Left																						

Table 1. Correlation coefficients (Pearson’s *r*) between each pair of conditions in Experiment 1. A color scale from blue to red reflects effect sizes from $r = -1$ to $r = 1$. Italics and bold font indicate significant results without ($\alpha = 0.05$) and with ($\alpha = 0.05/210$) Bonferroni correction, respectively. Binocular: binocular viewing condition; Right: monocular viewing condition with right eye; Left: monocular viewing condition with left eye.

corridor variant of the Ponzo illusion showed positive illusion magnitudes because the task was to adjust the disk that looks closer to the participant, unlike in the other variants of the Ponzo illusion (PZ, PZw, PZg), where the segment that looks further away (i.e., the top segment) was adjustable.

Correlations between pairs of conditions are reported in Table 1. In general, correlations were strong between different eye conditions of the same illusion and between different variants of an illusion (e.g., between the EB and EB2 conditions).

Mixed effects models

We tested for the effect of each predictor (eye condition, visual acuity, and sex) separately. First, and

most importantly, a likelihood ratio test showed that the eye condition did not significantly improve the model ($\chi^2(2) = 0.153, p = 0.926$). The eye condition was therefore removed from the model. Second, visual acuity did not significantly improve the model ($\chi^2(1) = 0.140, p = 0.710$) and was, hence, also removed. Last, sex did not significantly improve the model ($\chi^2(1) = 0.310, p = 0.579$). Hence, the best model did not include any predictor but only random effects for participants and for illusions. This model accounted for 82.6% of the variance in the data, which suggests that a large part of the variability in the data is accounted for by individual and between-illusion differences. Our results suggest that illusion magnitudes and individual differences in illusion magnitudes do not significantly vary with manipulation of the eye(s) to which the stimulus is presented and are not a function of visual acuity.

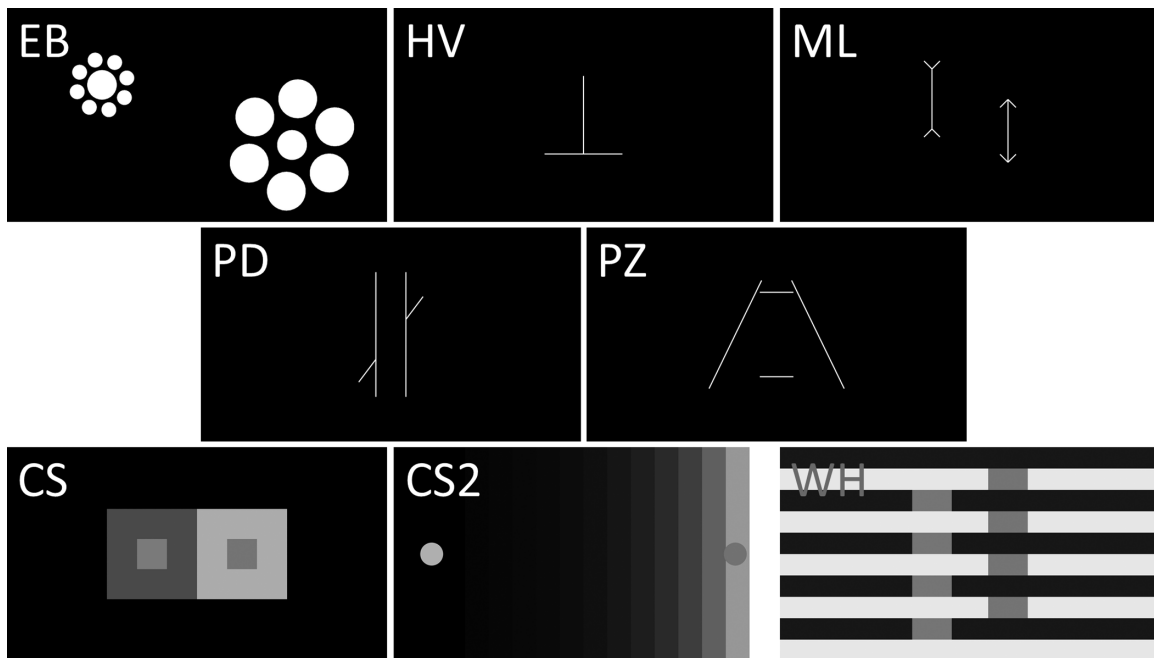


Figure 3. The eight illusions tested in [Experiment 2](#): Ebbinghaus (EB), horizontal-vertical (HV), Müller-Lyer (ML), Poggendorff (PD), Ponzo (PZ), two variants of the contrast (CS and CS2), and White (WH). Each illusion was tested with two reference-dependent conditions (for example, either the horizontal or vertical segment of the HV illusion was adjusted), making up 16 conditions. An adjustment procedure was used and there were two trials of each condition.

Experiment 2

Bistable percepts were shown to change over time ([Wexler, Duyck, & Mamassian, 2015](#)). Likewise, illusion decrement, that is, a decrease in the susceptibility to an illusion with repeated visual exposure, has been shown for a long time (e.g., [Coren & Girgus, 1972](#); [Judd, 1902](#); [Predebon, 2006](#)). Here, we wondered whether the individual differences in the perception of visual illusions vary over time.

Methods

Participants

Participants were 14 students of the Free University of Tbilisi, Georgia (7 females; mean age, 21 years; age range, 18–27 years). Participants signed informed consent before the experiment and were paid 10 GEL per hour. They were naïve to the purpose of the experiment. Procedures were conducted in accordance with the Declaration of Helsinki, except for preregistration (§ 35), and were approved by the local ethics committee.

Apparatus

The experiment was performed on a Windows-PC with LCD display (ASUS VG248QE; screen resolution:

1920 × 1080 pixels; refresh rate: 60 Hz). Stimuli were generated with Matlab (MathWorks Inc.) and the Psychophysics toolbox ([Brainard, 1997](#); [Pelli, 1997](#)).

Participants were seated at 60 cm from the screen and used a Logitech LS1 computer mouse for the adjustments. The experiment was conducted at the Free University of Tbilisi, Georgia.

Stimuli

Eight illusions were tested in [Experiment 2](#) ([Figure 3](#)): the Ebbinghaus (EB), horizontal-vertical (HV), Müller-Lyer (ML), Poggendorff (PD), Ponzo (PZ), two variants of the contrast (CS and CS2), and White (WH) illusions. As in [Experiment 1](#), a method of adjustment was used, i.e., participants had to adjust the size (EB, HV, ML, PZ), position (PD), or shade of gray (CS, CS2, WH) of an element to match the size, position, or shade of gray, respectively, of a reference on the screen by moving the computer mouse. The reference and adjustable elements were the central disks in the Ebbinghaus illusion, the horizontal and vertical segments in the horizontal-vertical illusion, the vertical segments with inward- and outward-pointing arrows in the Müller-Lyer illusion, the left and right parts of the interrupted diagonal in the Poggendorff illusion, the upper and lower horizontal segments in the Ponzo illusion, the inside squares in one variant of the contrast illusion (CS), the two disks in the other variant of the contrast illusion (CS2), and the two

columns of rectangles in the White illusion. There were two reference-dependent conditions for each illusion: one element (or series of elements, in the case of the White illusion) was in turn the reference or the adjustable element. For example, in the Ponzo illusion, the task was either to adjust the length of the upper horizontal segment so that it appeared to be the same length as the lower one, or to adjust the length of the lower horizontal segment so that it appeared to be the same length as the upper one. Stimuli were anti-aliased and lines were drawn with a width of about 0.1 degree. Illusions are described in further detail in the Supplementary File. Black and white had a luminance of $\approx 1 \text{ cd/m}^2$ and $\approx 176 \text{ cd/m}^2$, respectively.

Procedure

Participants were tested in 12 sessions with at least a 1-day break between two sessions. Participants were tested at days 1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 19, and 33 at about the same time of the day. Each session lasted for about 10 to 20 minutes.

On the first day of testing, the experimenter first explained the task to the participants. Then, participants completed eight practice trials (one of each illusion with a reference-dependent condition randomly chosen by the computer) in a random order to familiarize with the adjustment procedure.

Each session consisted of 32 trials since each condition (8 illusions \times 2 reference-dependent conditions = 16 conditions) was tested twice. The order of presentation of the 32 trials was randomly chosen by the computer for each participant and each session. Participants were asked to refrain from any prior knowledge about visual illusions. Importantly, no feedback was provided. The experimenter stayed in the experimental room during the experiment and answered any questions.

Data analysis

Intrarater reliabilities were computed as in [Experiment 1](#). As a measure of illusion magnitude for each participant, for both reference-dependent conditions and for each session, the adjustment from both trials were averaged into a mean adjustment, from which the reference was subtracted, similar to [Experiment 1](#). To test for the effect of reference-dependency, we computed Pearson correlations between the magnitudes of the two reference-dependent conditions of each illusion. Since the correlations were significant ($ps < 0.02$), we combined the two reference-dependent conditions of each illusion for further analyses. To estimate the effect of time (i.e., sessions) on the illusion magnitudes, we computed a repeated measures analysis of variance (ANOVA) with illusion and session as within-subject factors. A repeated

measures ANOVA catches any deviation from the mean. To examine how individual differences evolve with time (i.e., sessions), we computed Pearson correlations between pairs of sessions for each illusion and compared the correlation coefficients to simulated correlation coefficients.

We used both traditional null hypothesis significance testing (NHST) and Bayesian statistics. Unlike NHST, inferences about null results are allowed with Bayesian statistics. A Bayes factor (BF₁₀, later referred to as BF) smaller than 0.33 indicates evidence in favor of the null hypothesis, whereas a BF > 3 indicates substantial support for the alternative hypothesis ([Jeffreys, 1961](#)). BFs between 0.33 and 3 are considered inconclusive.

Results

Intrarater reliabilities

We computed ICCs for the two reference-dependent conditions of each illusion across sessions. All ICC coefficients were significant even when applying Bonferroni correction for multiple comparisons (see Supplementary Table S2). However, we like to mention that the correlation coefficients varied from medium (EB large: ICC coef. = 0.369) to large (EB small: ICC coef. = 0.815) according to [Cohen \(1988\)](#), but all suggesting large effects according to [Gignac and Szodorai \(2016\)](#). The illusion magnitudes of each participant averaged across sessions are shown for all conditions in Supplementary Figure S1.

Illusion magnitudes

[Figure 4](#) shows the illusion magnitudes for both reference-dependent conditions of each illusion (see also Supplementary Table S3). As expected, all illusions showed an over- and an under-adjusted condition. For example, the horizontal segment of the horizontal-vertical illusion was over-adjusted when compared to the vertical segment, and vice versa.

Are illusion magnitudes stable over time?

A repeated measures ANOVA was computed with illusion and session as within-subject factors. There was a significant interaction ($F(77, 1001) = 1.565, p = 0.002, \eta_p^2 = 0.107, \text{BF} = 3 \times 10^{270}$) and a significant main effect of illusion ($F(7, 91) = 28.14, p < 0.001, \eta_p^2 = 0.684, \text{BF} = 1 \times 10^{281}$), but the main effect of session was not significant ($F(11, 143) = 0.707, p = 0.73, \eta_p^2 = 0.052, \text{BF} = 7 \times 10^{-6}$). The data from the different illusions were further subjected to separate

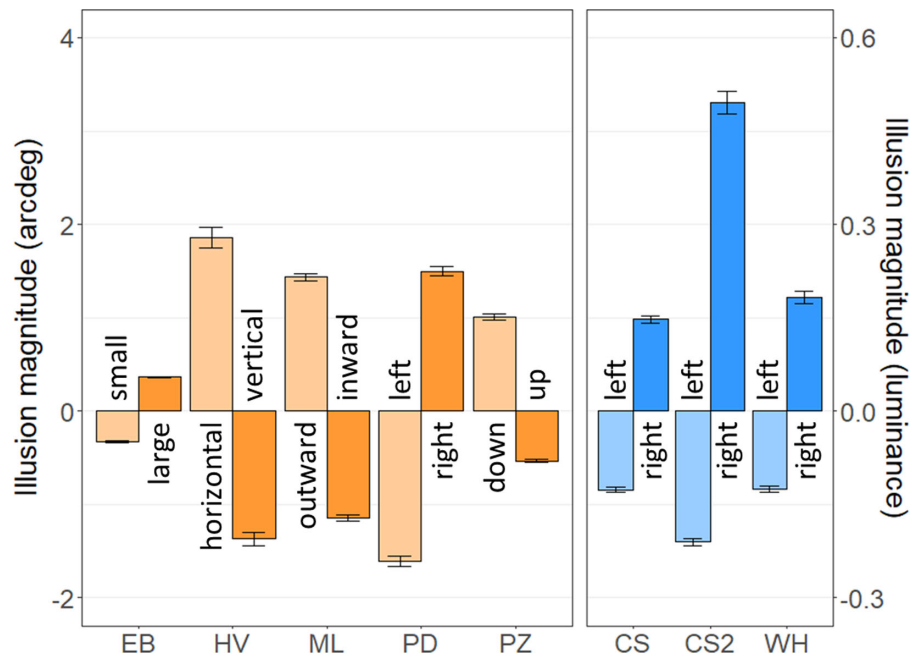


Figure 4. Illusion magnitudes \pm SEM (standard error of the mean) in Experiment 2 for both reference-dependent conditions of each illusion. Positive and negative magnitudes indicate over- and under-adjustments compared to the reference, respectively. Left panel (orange): illusions with adjustment of size, length, or position measured in units of arcdeg; right panel (blue): illusions with contrast adjustment measured in normalized luminance.

one-way repeated measures ANOVAs with session as within-subject factor, which revealed a significant main effect of session in the Ebbinghaus illusion and in one variant of the contrast (CS2) illusion (EB: $F(11,143) = 1.893$, $p = 0.045$, $\eta_p^2 = 0.127$, BF = 0.788; CS2: $F(11,143) = 2.679$, $p = 0.004$, $\eta_p^2 = 0.171$, BF = 7.939) but not in the other illusions (HV: $F(11,143) = 1.687$, $p = 0.082$, $\eta_p^2 = 0.115$, BF = 0.435; ML: $F(11,143) = 1.034$, $p = 0.420$, $\eta_p^2 = 0.074$, BF = 0.073; PD: $F(11,143) = 1.452$, $p = 0.156$, $\eta_p^2 = 0.100$, BF = 0.226; PZ: $F(11,143) = 0.895$, $p = 0.547$, $\eta_p^2 = 0.064$, BF = 0.050; CS: $F(11,143) = 0.780$, $p = 0.659$, $\eta_p^2 = 0.057$, BF = 0.037; WH: $F(11,143) = 1.473$, $p = 0.148$, $\eta_p^2 = 0.102$, BF = 0.240). Note that BF's were inconclusive in the case of the Ebbinghaus and horizontal-vertical illusions. Hence, there were differences in the mean magnitudes of the Ebbinghaus and one variant of the contrast illusions across sessions (Figure 5), with a general increase in illusion magnitudes over time. However, this does not imply that individual differences are not stable. For example, it may be that all participants undergo a similar change in the susceptibility to one illusion over time (e.g., a 20% increase within a certain amount of time). In this case, individual differences would remain stable, which would show up as strong and significant between-session correlations for each illusion.

Are individual differences in the perception of illusion magnitudes stable over time?

To examine how individual differences evolve with time, we computed between-session correlations for each illusion (Figure 6). Correlations were strong in general and tended to weaken between pairs of sessions that were further apart in time (especially in the Müller-Lyer and Pöschel-Zeigler illusion).

Then, we wondered whether such patterns of correlations may result from stationary time series, i.e., stable individual differences over time. For this purpose, we simulated the individual illusion magnitudes from normal distributions centered on the behavioral data. In other words, for each participant and each illusion, we computed the mean magnitude and standard deviation across sessions and randomly picked 12 values (for the 12 sessions) from a normal distribution centered and scaled on these values. Between-session correlations were then computed for each illusion and averaged across 10,000 simulations. We show the behavioral and simulated correlation coefficients in Figure 7 as a function of the time-lag (i.e., the time difference in days) between each pair of sessions.

If there were a time effect on the individual differences in the perception of visual illusions, correlations from simulated data should be stronger than correlations from behavioral data and the difference may strengthen

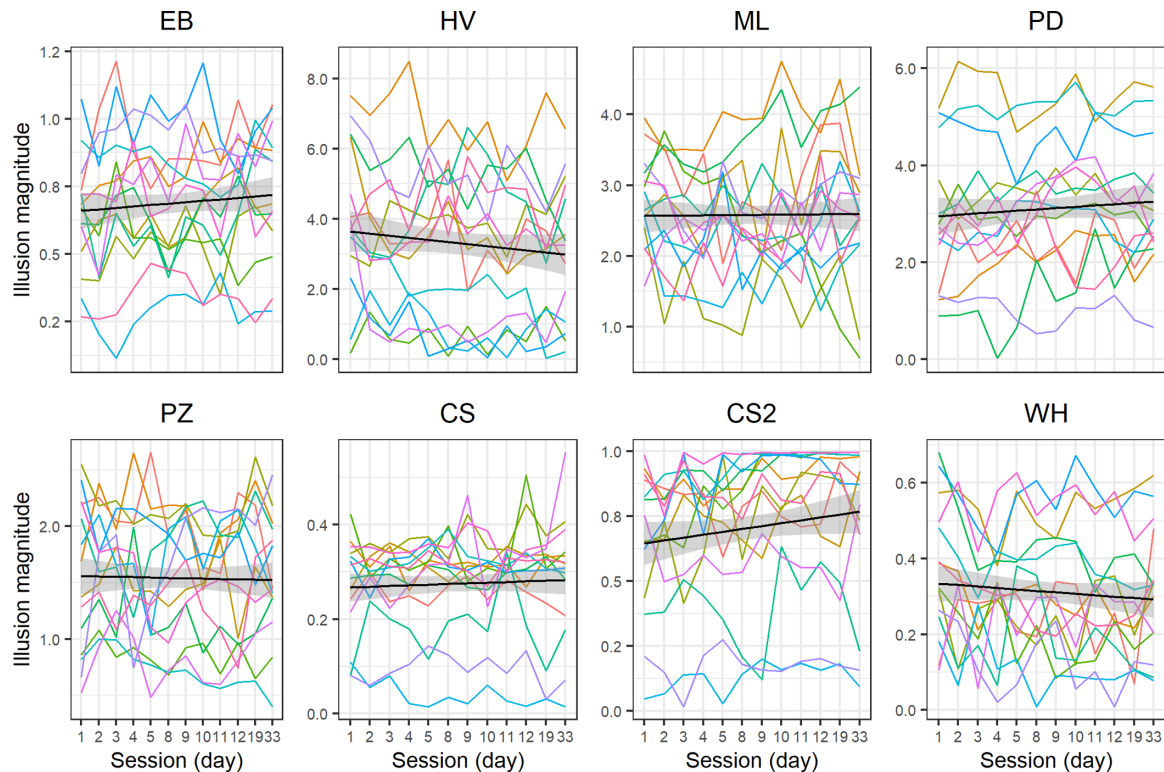


Figure 5. Illusion magnitudes plotted for each illusion (sub-panels) tested in [Experiment 2](#) across sessions. Colors represent individuals. A linear regression (black line) was fitted with 95% confidence interval (shaded area). Illusion magnitudes are plotted in units of arcdeg for the EB, HV, ML, PD, and PZ illusions and in normalized luminance for contrast illusions (CS, CS2, WH).

with the time-lag (i.e., when two sessions are farther away in time). Paired two-tailed t -tests were computed between the behavioral and simulated correlation coefficients for each illusion. All of them resulted in non-significant differences with BF smaller than 0.33 (EB: $t(65) = 0.796$, $p = 0.429$, $d = 0.098$, BF = 0.183; HV: $t(65) = -0.036$, $p = 0.971$, $d = 0.004$, BF = 0.135; ML: $t(65) = -0.461$, $p = 0.647$, $d = 0.057$, BF = 0.149; PD: $t(65) = 0.069$, $p = 0.945$, $d = 0.009$, BF = 0.135; PZ: $t(65) = -0.303$, $p = 0.763$, $d = 0.037$, BF = 0.141; CS: $t(65) = 0.258$, $p = 0.797$, $d = 0.032$, BF = 0.139; WH: $t(65) = -0.283$, $p = 0.778$, $d = 0.035$, BF = 0.140), except for the second variant of the contrast illusion, which showed inconclusive BF (CS2: $t(65) = 1.879$, $p = 0.065$, $d = 0.231$, BF = 0.705). Overall, these results suggest that individual differences in visual illusions are stable across sessions.

Experiment 3

We previously measured the susceptibility to visual illusions with an adjustment procedure ([Cretenoud et al., 2019](#); [Cretenoud, Francis, et al., 2020](#); [Cretenoud, Grzeczowski, et al., 2020](#); [Grzeczowski et al., 2017](#);

[Grzeczowski et al., 2018](#)). However, forced choice response modalities, for example, a method of constant stimuli, are usually thought to be more reliable (e.g., [Todorović & Jovanović, 2018](#)). Here, we compared the individual differences when two visual illusions were tested with an adjustment procedure and a method of constant stimuli.

Methods

Participants

Twenty students from the EPFL, Switzerland, and Université de Lausanne were tested (seven females; mean age, 23 years; age range, 19–28 years). None of the participants had taken part in [Experiments 1 or 2](#), and all were naïve to the purpose of the experiment. Participants signed informed consent before the experiment and were paid 20 Swiss Francs per hour and an extra amount of 5 Swiss Francs to compensate for the constraints related to Covid-19. Procedures were conducted in accordance with the Declaration of Helsinki, except for preregistration (§ 35) and were approved by the local ethics committee.

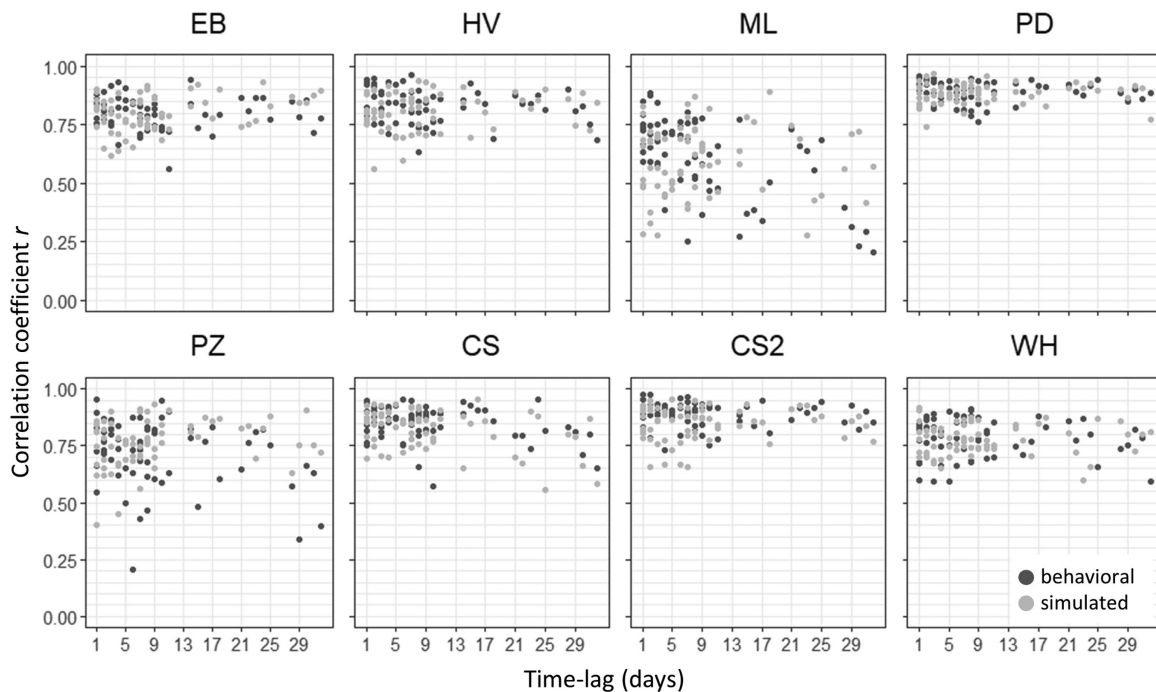


Figure 7. Correlation coefficients (r) as a function of the time-lag (i.e., the time difference) between sessions in days from behavioral data (in dark grey) and simulated data (in light grey) for each illusion in Experiment 2. Correlations from simulated data were expected to be stronger than correlations from behavioral data under the hypothesis that individual differences in visual illusions vary with time. However, we did not observe any significant differences between simulated and behavioral correlation coefficients.

Power analysis

We ran Bayesian simulations as a power analysis to determine the sample size needed for this experiment. Assuming a small effect size of $d = 0.1$, a Bayes factor smaller than 0.33 was computed for $n \geq 20$ (averaged across 1000 simulations). Hence, 20 participants were tested. A sensitivity analysis revealed that an effect size of $r = 0.552$ could be detected with $n = 20$, $\alpha = 0.05$, and 80% power.

Apparatus

Stimuli were presented on a BenQ XL2540 LCD monitor (screen resolution: 1920×1080 pixels, refresh rate: 60 Hz) using Matlab (MathWorks Inc.) and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). The distance to the screen was approximately 60 cm. Participants used a Logitech M-BJ58 computer mouse for the adjustments. The experiment was conducted in the Laboratory of Psychophysics at EPFL, Switzerland.

Stimuli

Two illusions were tested: the Ebbinghaus (EB) and Müller-Lyer (ML) illusions (the metrics were the same

as in Experiment 2 except when mentioned; see Figure 3 and the Supplementary File).

Procedure

Both illusions were tested with an adjustment procedure, as in Experiments 1 and 2, and with a method of constant stimuli. In the EB large condition, the central disk surrounded by large flankers was the target, whereas the central disk surrounded by small flankers was the reference. In the EB small condition, the target and the reference were reversed. Similarly, in the Müller-Lyer illusion, the target was either the vertical segment with inward-pointing arrows (ML inward condition) or the one with outward-pointing arrows (ML outward condition). Hence, one element was in turn the target or reference, that is, each illusion was tested with two reference-dependent conditions.

For the adjustment procedure, each illusion and reference-dependent condition was tested twice (2 illusions \times 2 reference-dependent conditions \times 2 trials = 8). For example, in the Müller-Lyer illusion, the vertical segment with inward-pointing arrows was adjusted twice, and similarly for the vertical segment with outward-pointing arrows. In addition, each illusion was tested twice without flankers (EB) or

arrows (ML), that is, as control conditions (2 illusions \times 2 trials = 4).

Participants were asked to adjust the size of the target to match the size of the reference by moving the computer mouse on the vertical axis. At the beginning of each trial, the diameter (EB) and length (ML) of the target was randomly set in the range between 0° and 6.7° and between 2.1° and 19.9° , respectively.

For the method of constant stimuli, participants had to report whether the target was larger or smaller than the reference by using the up or down key on the keyboard, respectively. The target in each illusion and reference-dependent condition was tested 20 times with eight different increments (i.e., eight sizes; 2 illusions \times 2 reference-dependent conditions \times 8 increments \times 20 trials = 640). For each illusion and each reference-dependent condition, the incremental range was centered on the mean adjustment, i.e., the point of subjective equality (PSE), from Experiment 2 and covered three times the absolute difference between the mean adjustment from Experiment 2 and the reference (see Supplementary Table S4). If a participant shows an illusion bias similar to the average bias observed in Experiment 2, 50% of all responses related to that illusion are expected to be “larger.” Similarly, 50% of the responses are expected to be “larger” if a participant has no illusory bias.

In addition, control conditions were tested with eight increments equally distributed around the reference value and covering the range between the mean adjustments for both reference-dependent conditions in Experiment 2 (2 illusions \times 8 increments \times 20 trials = 320; see Supplementary Table S4).

For both methods, the position of the target and reference on the screen was counterbalanced, that is, the reference was displayed in the right half-screen in half of the trials and in the left half-screen in the other half of the trials. Moreover, the position of the target and reference along the y-axis was randomized with the screen size as limits and with the constraint that the target and reference were never at the same position along the y-axis to avoid any direct horizontal comparison between them. For the method of constant stimuli, a red square cue (0.5° in side, vertically centered) displayed on the left or right of the stimulus indicated which element was the target.

First, the experimenter explained the task to the participants. There were four warming up trials (2 illusions \times 2 methods). Then, the experiment was split into eight blocks: each illusion (EB, ML) was tested with both methods (adjustment, method of constant stimuli) and with and without context. The order of the eight blocks and the order of the trials within a block were randomized across participants. We asked participants to ignore any prior knowledge about illusions. Stimuli were shown until a response was given, that is, a mouse click with the

adjustment procedure or a key press with the method of constant stimuli. There was no time restriction and no feedback.

Data analysis

As in Experiments 1 and 2, the illusion magnitudes are expressed as a size difference compared to the reference with positive and negative values indicating over- and under-adjustments, respectively.

As in Experiments 1 and 2, we computed intrarater reliabilities for each condition in the method of adjustment and then averaged both trials of each condition into a mean adjustment.

Psychometric curves were fitted for each condition with the method of constant stimuli to determine the PSE, that is, the size of the target needed so that the participant perceived both the target and reference to be the same size (2 illusions \times [2 reference-dependent conditions + 1 control] = 6 conditions). To this end, we summed the reports perceived as “larger” than the reference at each increment. Using a cumulative Gaussian function with a lapse rate of 0.02, we then defined the PSE as the size of the target corresponding to 50% of “larger” responses. We screened the data for invalid psychometric fits, that is, no PSE was extracted when the mu or sigma parameter of the underlying Gaussian function was at the edge or outside of the search space, which was defined as a function of the incremental range. Based on that criterion, 2.5% of the PSEs were missing.

To check for outliers, the adjustment trials and the PSEs from the psychometric function for each condition (six conditions) were then standardized by computing modified z-scores, which are more robust than the commonly used z-scores because it makes use of the median and median absolute deviation instead of the mean and standard deviation, respectively. As suggested by Iglewicz and Hoaglin (1993), modified z-scores larger than 3.5 were considered as outliers. Based on that criterion, 4.2 and 1.7% of the data with the adjustment procedure and method of constant stimuli, respectively, was removed for further analyses. Outlying and missing data points were imputed using the “mice” function from the mice R package with method “norm” (Bayesian linear regression with 20 imputation samples).

To compare the illusion magnitudes from the adjustment procedure and the method of constant stimuli, we first accounted for the bias in the control conditions. Hence, for each illusion and each method, we subtracted the control condition from both reference-dependent conditions, for example, we subtracted the EB control condition from the EB large and EB small conditions in the adjustment method, and similarly in the method of constant stimuli. Then, we computed paired *t*-tests and Pearson correlations

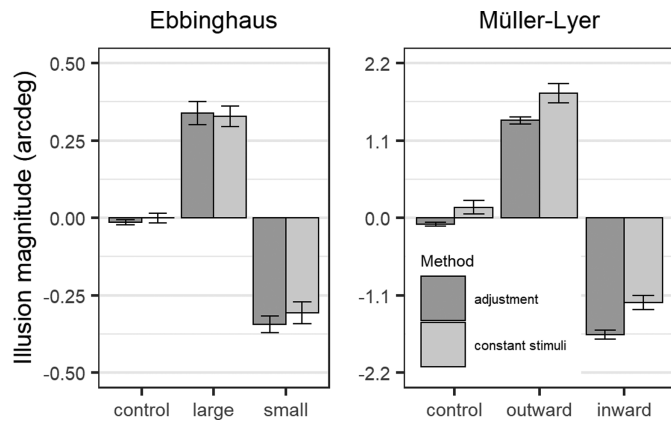


Figure 8. Illusion magnitudes \pm SEM (standard error of the mean) in [Experiment 3](#) for the control condition and both reference-dependent conditions of each illusion (left panel: Ebbinghaus illusion; right panel: Müller-Lyer illusion). Positive and negative magnitudes indicate over- and under-adjustments compared to the reference, respectively. Dark grey: adjustment procedure; light grey: method of constant stimuli.

between the illusion magnitudes from both methods. As in [Experiment 2](#), we both used traditional NHST and Bayesian statistics.

Results

Intrarater reliabilities

The noncontrol conditions showed significant intraclass correlations with medium to large effect sizes, according to [Gignac and Szodorai \(2016\)](#); EB large: ICC coef. = 0.550, 95% CI [0.227, 0.764], $p = 0.005$; EB small: ICC coef. = 0.554, 95% CI [0.233, 0.766], $p = 0.005$; ML outward: ICC coef. = 0.420, 95% CI [0.060, 0.683], $p = 0.029$, except for the ML inward condition, which only showed a small and non-significant effect size (ICC coef. = 0.130, 95% CI [-0.251, 0.476], $p = 0.287$). The control conditions of both illusions showed weaker intraclass correlations (EB control: ICC coef. = 0.038, 95% CI [-0.335, 0.401], $p = 0.434$; ML control: ICC coef. = 0.377, 95% CI [0.010, 0.655], $p = 0.046$), which suggests that there is no strong measurement bias and that the residual bias is mainly due to noise.

Illusion magnitudes

[Figure 8](#) shows the illusion magnitudes for all conditions (for more details, see [Supplementary Table S5](#)). Illusion magnitudes were approximately in the same range as in previous experiments (e.g., [Experiment 2](#); see also [Cretenoud, Francis, et al., 2020](#); [Cretenoud, Grzeczowski, et al., 2020](#); [Cretenoud et al., 2019](#)). As expected, the control conditions led to very weak effects.

		EB large	EB small	ML inward	ML outward
Paired <i>t</i> -tests	<i>t</i>	-0.953	0.617	1.648	1.662
	<i>p</i>	0.353	0.544	0.116	0.113
	BF	0.347	0.276	0.735	0.748
Correlations	<i>r</i>	0.602	0.671	0.324	0.615
	<i>p</i>	0.005	0.001	0.164	0.004
	BF	10.355	27.743	1.021	12.200

Table 2. Statistics from paired *t*-tests and correlations between the measures of illusion magnitudes from the adjustment procedure and method of constant stimuli in [Experiment 3](#), after accounting for the bias in the respective control conditions.

Comparing both methods

We accounted for the control bias by subtracting the control condition from the two reference-dependent conditions for each illusion in each method. Then, we computed paired *t*-tests to compare the means between the measures of illusion magnitudes from the adjustment procedure and from the method of constant stimuli. Results are reported in [Table 2](#). None of the four tests revealed a significant difference. We also computed Bayesian paired *t*-tests, which resulted in Bayes factors in the range of 0.276 to 0.748. There was substantial support for the null hypothesis, that is, both methods led to similar illusion magnitudes, in the EB small condition. The EB large, ML inward and ML outward conditions were associated with an inconclusive BF.

Correlations between the measures of illusion magnitudes from both methods resulted in large and significant effect sizes for the EB large, EB small, and ML outward conditions ([Table 2](#) and [Figure 9](#)). However, the ML inward condition resulted in a smaller and non-significant effect size of $r = 0.324$ with BF = 1.021.

Discussion

There are large individual differences in visual illusions. Here, we examined how stable individual differences in the perception of visual illusions are. First, we observed that illusory percepts are reliable interocularly and are not a function of visual acuity. Second, individual differences in the perception of visual illusions were reliable over time. Third, the methods of adjustment and constant stimuli showed reliable individual differences, except for one condition. Hence, our results suggest that the mixed results about factors for visual illusions, i.e., there is a unique common or several specific factors underlying illusions,

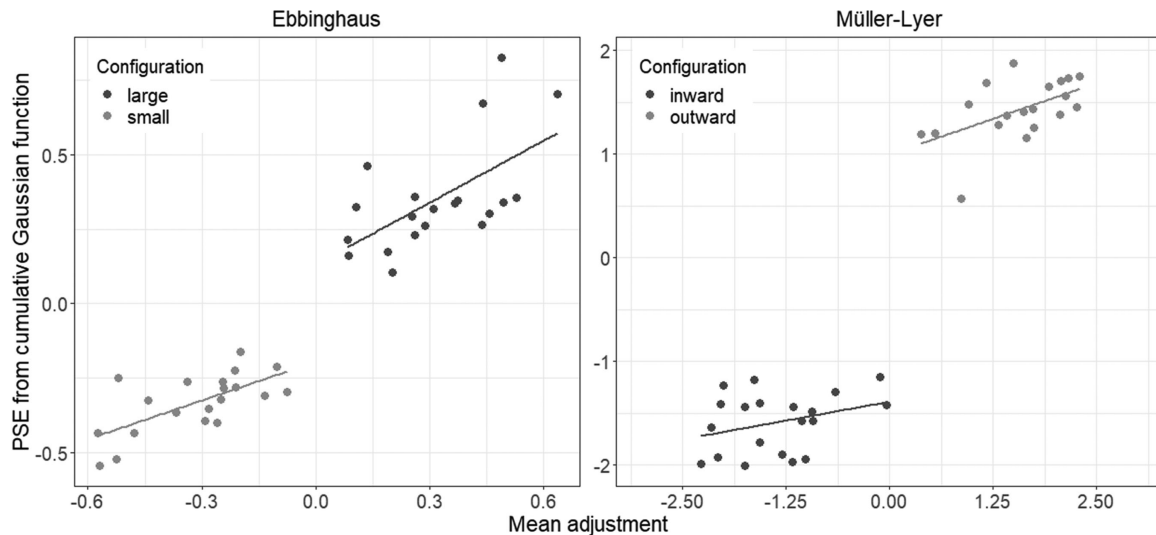


Figure 9. Correlations (i.e., linear regressions) between the measures from the adjustment procedure and method of constant stimuli in units of arcdeg for both reference-dependent conditions of the Ebbinghaus (left panel) and Müller-Lyer (right panel) illusions in Experiment 3, after accounting for the bias in the respective control conditions.

do not result from unstable individual differences in the perception of visual illusions across eyes, time, and measurement methods.

Recent studies suggested a multifactorial structure underlying vision, such as in oculomotor tasks (Bargary et al., 2017), bistable paradigms (Brascamp et al., 2018; Cao, Wang, Sun, Engel, & He, 2018; Wexler et al., 2015), local and global visual processing (Chamberlain et al., 2017), and in the use of expectations and knowledge priors (Tulver, Aru, Rutiku, & Bachmann, 2019), which argues against a unique, general factor for vision as proposed previously (Halpern, Andrews, & Purves, 1999). Similarly, we previously claimed that there are illusion-specific factors (e.g., Cretienoud et al., 2019). For example, we observed strong correlations between different variants of the Ebbinghaus illusion, which differed in color, size, or shape, suggesting that there is a factor specific to the Ebbinghaus illusion. Unlike perceptual learning, which is orientation-specific, illusion magnitudes strongly correlated for different orientations.

Hence, we expected illusion magnitudes to be independent of eye and visual acuity. The results in Experiment 1 further support that claim as we observed no significant differences in the illusion magnitudes between monocular (left vs. right) and binocular viewing conditions. Similarly, visual acuity did not show any significant effect on the illusion magnitudes. Note that in Experiment 1 the two trials of each condition were presented sequentially, that is, one after the other, which may inflate the intrarater reliabilities. However, the size of the adjustable target was randomly set at the beginning of each trial, which makes two trials (even presented one after the other) hardly comparable.

Stable individual differences were recently reported in the perception of different variants of the Ponzo illusion, which differed in context, e.g., with line-drawing or real-world perspective (Cretienoud, Grzeczowski, et al., 2020), suggesting that similar mechanisms are at hand when geometric or real-world depth cues are presented (but see Leibowitz et al., 1969; Wagner, 1977). In Experiment 1, we similarly observed strong correlations between the magnitudes of the different variants of the Ponzo illusion. Note, however, that the correlations between the corridor variant (PZc) and the three other variants of the Ponzo illusion (PZ, PZw, and PZg) were weaker than between other pairs of Ponzo variants, especially in the monocular conditions (Table 1). In addition, there were strong correlations between the susceptibilities to the corridor variant of the Ponzo illusion (PZc) and the Ebbinghaus illusion (EB). Full and partial interocular transfer were previously observed in the Ponzo and Ebbinghaus illusions, respectively, suggesting that different mechanisms underlie the two illusions (Song, Schwarzkopf, & Rees, 2011). However, partial interocular transfer was observed in the effect of linear cues (but full transfer for texture gradients) in the perception of the corridor illusion (Yildiz, Sperandio, Kettle, & Chouinard, 2021), which suggests that the perceptual rescaling depends on the nature of the cues (i.e., lower- or higher-order features). Hence, in the present investigation, it may be that the corridor and the Ebbinghaus illusions rely on more similar features than the corridor and other variants of the Ponzo illusion.

For more than a century, the susceptibility to visual illusions was suggested to decrease with repeated visual exposures, which was called the illusion decrement (e.g.,

Coren & Girgus, 1972; Judd, 1902; Predebon, 2006). For example, a Müller-Lyer illusion decrement was observed under different viewing conditions, such as free viewing (e.g., Festinger, White, & Allyn, 1968) or at fixation (e.g., Day, 1962). In [Experiment 2](#), there were no significant differences in the mean illusion magnitudes across time, except for the Ebbinghaus and a variant of the contrast illusion. Importantly, we do not claim that the illusion magnitudes are stable over time. Indeed, it may be that there are cumulative changes over time, as observed in paradigms with bistable stimuli (Wexler, 2018; Wexler et al., 2015). Instead, we wondered whether individual differences in the perception of visual illusions are stable. Hence, we were not interested in the changes in illusion magnitudes per se, but rather in comparing these changes across individuals over time. For example, we wondered whether a participant, who showed a stronger susceptibility to the Ebbinghaus illusion compared to other participants in the first testing session, showed this stronger susceptibility also a month later. We indeed found that this is the case, even though we do not preclude that individual differences in the perception of visual illusions may slightly vary over time. It seems that individual differences are largely stable over time, while illusion magnitudes are not. Note that we did not provide feedback to prevent learning. In addition, each illusion was tested only four times in each testing session (2 reference-dependent conditions \times 2 trials), resulting in a rather short exposure to the stimuli.

In [Experiment 2](#), the magnitudes of one variant of the Ebbinghaus illusion (EB2) were weaker compared to the other variant of the same illusion (EB), which may come from the close proximity of the flankers compared to the adjustable target (i.e., the size of the adjustable target could not be much increased without overlapping with the flankers). Despite this limitation, we observed rather strong correlations between both variants of the Ebbinghaus illusion, suggesting that the EB2 magnitudes are reliable. Interestingly, the magnitudes of the White and both contrast (CS and CS2) illusions strongly—but not always significantly—correlated (e.g., at d9 – CS-WH: $r = 0.473$, $p = 0.088$; CS2-WH: $r = 0.491$, $p = 0.074$; at d10 – CS-WH: $r = 0.615$, $p = 0.019$; CS2-WH: $r = 0.638$, $p = 0.014$), even though the White illusion is phenomenologically different compared to the contrast illusions.

Different methods of illusion measurements have been compared in the past. For example, [Coren and Girgus \(1972\)](#) compared five methods: an adjustment procedure, a method of reproduction, a selection from a graded series, and two subjective methods, that is, rating scale and magnitude estimation. Unlike the subjective methods, the other three methods showed expected and significant effects in the Ebbinghaus and Müller-Lyer illusions. For example, the inward-pointing Müller-Lyer illusion magnitude was shown to significantly decrease

with increasing angle of the fins. Similarly, the apparent size of the Ebbinghaus target significantly decreased when surrounding flankers increased in size. Nevertheless, the authors did not compute correlations between the illusion magnitudes from these different methods. The aim of [Experiment 3](#) was to compare the individual differences in illusion magnitudes (rather than the illusion magnitudes per se) from two methods, namely an adjustment method and a method of constant stimuli. The illusion magnitudes measured with both methods were strongly correlated (with large associated BFs) in three out of four conditions. Only the ML inward condition did not show a significant correlation (BF = 1.021; inconclusive). Overall, the results suggest that individual differences are largely reliable across both methods. Similarly, [Schwarzkopf and colleagues \(2011\)](#) observed strong correlations between an adjustment procedure (with eight trials per condition) and a method of constant stimuli in the estimation of the susceptibility to the Ebbinghaus and Ponzo corridor illusions.

Most illusion studies used forced choice responses (see [King, Hodgekins, Chouinard, Chouinard, & Sperandio, 2017](#)), such as the method of constant stimuli used in [Experiment 3](#). However, there are several weaknesses associated with that method. First, the method of constant stimuli requires a large number of trials. For example, in [Experiment 3](#), there were 960 versus 12 trials in the method of constant stimuli versus adjustment procedure, respectively. Repeated exposures may lead to fatigue or learning effects, as mentioned earlier. Note that in [Experiment 3](#), the cue was presented together with the stimulus to decrease the proportion of trials answered wrongly (i.e., the lapse rate) because of a lack of attention following the large number of trials in the method of constant stimuli.

Illusions are a matter of perceptual bias. In addition, illusion magnitudes may be contaminated by decisional bias, for example in the method of constant stimuli. Indeed, individual differences in the illusion magnitudes observed with that method may not only reflect a difference in the sensitivity to visual illusions but also in the participants' subjective, decisional criterion. For example, a participant may always report that the reference stimulus is larger when he or she is unsure. This type of decisional bias is weakened in the adjustment procedure because participants are not asked to pick one out of two elements (target or reference). Note, however, that we do not claim that the adjustment procedure is completely free from response bias. For example, participants may use different strategies when they have to adjust an element that is obviously larger versus smaller than the reference.

Decisional biases may especially affect the validity of conclusions made when comparing a clinical population to a control group. Using a 2-AFC roving pedestal method, which reduces decisional bias (there

is no clue about which element is the reference), Manning, Morgan, Allen, and Pellicano (2017) reported no substantial difference in the susceptibility to the Ebbinghaus and Müller-Lyer between autistic and typically developing children. However, using an adjustment procedure, the authors observed a weak evidence in favor of a group difference in the Müller-Lyer illusion (but not in the Ebbinghaus illusion), with autistic children showing larger illusion magnitudes compared to typically developing children. The authors suggested that the adjustment procedure is susceptible to atypical decisional bias, for example in clinical populations. In addition, they claimed that the Müller-Lyer illusion is more prone to decisional bias compared to the Ebbinghaus illusion, which may explain why we observed a weaker correlation between the Müller-Lyer inward illusion magnitudes when tested with an adjustment procedure and a method of constant stimuli (Experiment 3). However, Manning and colleagues (2017) did not test the method of constant stimuli (2-AFC) as we did. Note that negative relationships between autistic-like traits and susceptibility to visual illusions were previously reported (e.g., Chouinard, Noulty, Sperandio, & Landry, 2013; Happé, 1996; but see Chouinard, Unwin, Landry, & Sperandio, 2016), whereas similar patterns of between-illusion correlations were observed in patients with schizophrenia and healthy controls (Grzeczowski et al., 2018). In contrast, the personality dimension of schizotypy was shown to correlate with the likelihood of an individual to see meaning in a noisy, meaningless image (Partos, Cropper, & Rawlings, 2016).

More than being playful (e.g., when testing children), the adjustment procedure as used here and previously (Cretenoud et al., 2019; Cretenoud, Francis, et al., 2020; Cretenoud, Grzeczowski, et al., 2020; Grzeczowski et al., 2017) also has weaknesses. For example, it may be argued that two trials of each condition are too few. Indeed, a mis-click or poor attention paid to one of the trials biases the average, thus leading to a poor estimate of the illusion magnitude. However, intrarater reliabilities were in general significant in the adjustment procedure. In addition, the results from Experiment 3 showed that individual differences in the perception of visual illusions are in general similar across both methods, emphasizing that an adjustment procedure is amenable to replace a method of constant stimuli when testing healthy participants. Note that both methods were non-speeded, i.e., the stimuli were shown until a response was given.

In previous publications, we found that there are only weak correlations between the susceptibility to different illusions. For example, out of 15 between-illusion correlations, only one was significant when 113 participants were tested (Grzeczowski et al., 2017). Similarly, we previously reported only 65 out of 720 significant between-illusion correlations (Experiment

2 in Cretenoud et al., 2019). Effect sizes were in most cases rather small (e.g., correlation coefficients r ranged between -0.12 and 0.23 in Grzeczowski et al., 2017). These results are surprising because in the latter publication, all illusions were spatial illusions, which are often implicitly or explicitly assumed to share the same mechanism (e.g., Coren et al., 1976; Ninio, 2014; Piaget, 1961). However, a shared mechanism should have led to substantial correlations. The weak correlations cannot be explained by large intraobserver variability because intrarater reliabilities were mostly significant, as in the present study. In addition, in Cretenoud et al. (2019), we found large within-illusion correlations for different luminance conditions including isoluminant ones, for 19 different variants of the Ebbinghaus illusion including two conditions with rotating flankers, and when illusions were tested under different orientations. Finally, individual differences were stable when illusions were presented as line drawings or within natural scenes like train tracks (Cretenoud, Grzeczowski, et al., 2020). It seems that each illusion makes up its own factor.

Here, we showed that individual differences in a monocular and binocular viewing conditions are robust, showing that differences between eyes are of little relevance. Hence, the mechanisms underlying illusions seem to occur after binocular fusion. In addition, we backed up our previous results by showing that individual differences in the perception of visual illusions are independent of the methods used and over time. Our results provoke the questions to what extent taxonomies of illusions are useful and what we can learn from detailed analysis of mechanisms underlying visual illusions, since these mechanisms seem to be rather idiosyncratic.

Our findings about illusions nicely fit into a larger picture about common factors for vision in general because there are not only many factors for illusions but also for many visual paradigms. For example, individual differences in contrast sensitivity (Peterzell, 2016; Peterzell, Scheffrin, Tregear, & Werner, 2000), hue scaling (e.g., Emery et al., 2017a; Emery et al., 2017b), color matching (Webster & MacLeod, 1988), stereopsis (Peterzell, Serrano-Pedraza, Widdall, & Read, 2017), and in the effects of priors (Tulver et al., 2019) were suggested to rely on several specific factors. All these studies clearly argue against the widespread intuition that there are only a few mechanisms behind vision, which just need to be unearthed. It rather seems that we are dealing with a plethora of idiosyncratic mechanisms, which provokes the question to what extent the study of detailed mechanisms can lead to a unified theory of visual processing.

To summarize, we showed that the individual differences in the perception of visual illusions are reliable interocularly (Experiment 1), over time (Experiment 2), and when measured with an adjustment

procedure or a method of constant stimuli (Experiment 3). Hence, the mixed results previously reported, i.e., there is a unique or several specific common factors for illusions, are unlikely related to unstable individual differences across eyes, time, and measurement methods. Future studies may examine the effect of other differences in the experimental design of the studies, such as speeded versus non-speeded tasks, and in the statistical methods used to determine the number of factors to extract.

Keywords: individual differences, illusions, reliability, measurement method

Acknowledgments

The authors thank Marc Reppow for technical support and Pornbhussorn Kanchanakanok for helping with data collection (Experiment 1).

Supported by the project “Basics of visual processing: from elements to figures” (project no. 320030_176153/1) of the Swiss National Science Foundation (SNSF) and by a National Centre of Competence in Research (NCCR Synapsy) grant from the SNSF (51NF40-185897).

Commercial relationships: none.

Corresponding author: Aline F. Cretenoud.

Email: aline.cretenoud@gmail.com.

Address: Ecole Polytechnique Fédérale de Lausanne (EPFL), Laboratory of Psychophysics, Brain Mind Institute, 1015 Lausanne, Switzerland.

References

- Aftanas, M. S., & Royce, J. R. (1969). A factor analysis of brain damage tests administered to normal subjects with factor score comparisons across ages. *Multivariate Behavioral Research*, 4(4), 459–481.
- Bach, M. (1996). The Freiburg Visual Acuity test automatic measurement of visual acuity. *Optometry and Vision Science*, 73, 49–53, <http://dx.doi.org/10.1097/00006324-199601000-00008>.
- Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging*, 12(1), 12–21, <https://doi.org/10.1037/0882-7974.12.1.12>.
- Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. D. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Research*, 141, 157–169, <https://doi.org/10.1016/j.visres.2017.03.001>.
- Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, 19(1), 3–11, <https://doi.org/10.2466/pr0.1966.19.1.3>.
- Bosten, J. M., Goodbourn, P. T., Bargary, G., Verhallen, R., Lawrance-Owen, A. J., Hogg, R. E., . . . Mollon, J. D. (2017). An exploratory factor analysis of visual performance in a large population. *Vision Research*, 141, 303–316, <https://doi.org/10.1016/j.visres.2017.02.005>.
- Brainard, D. H. (1997). The Psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Brascamp, J. W., Becker, M. W., & Hambrick, D. Z. (2018). Revisiting individual differences in the time course of binocular rivalry. *Journal of Vision*, 18(7), 1–20, <https://doi.org/10.1167/18.7.3>.
- Cao, T., Wang, L., Sun, Z., Engel, S. A., & He, S. (2018). The independent and shared mechanisms of intrinsic brain dynamics: insights from bistable perception. *Frontiers in Psychology*, 9:589, 1–11, <https://doi.org/10.3389/fpsyg.2018.00589>.
- Cappe, C., Clarke, A., Mohr, C., & Herzog, M. H. (2014). Is there a common factor for vision? *Journal of Vision*, 14(8), 1–11, <https://doi.org/10.1167/14.8.4>.
- Chamberlain, R., Van der Hallen, R., Huygelier, H., Van de Cruys, S., & Wagemans, J. (2017). Local-global processing bias is not a unitary individual difference in visual processing. *Vision Research*, 141, 247–257, <https://doi.org/10.1016/j.visres.2017.01.008>.
- Chouinard, P. A., Noulty, W. A., Sperandio, I., & Landry, O. (2013). Global processing during the Müller-Lyer illusion is distinctively affected by the degree of autistic traits in the typical population. *Experimental Brain Research*, 230(2), 219–231, <https://doi.org/10.1007/s00221-013-3646-6>.
- Chouinard, P. A., Unwin, K. L., Landry, O., & Sperandio, I. (2016). Susceptibility to Optical Illusions Varies as a Function of the Autism-Spectrum Quotient but not in Ways Predicted by Local–Global Biases. *Journal of Autism and Developmental Disorders*, 46(6), 2224–2239, <https://doi.org/10.1007/s10803-016-2753-1>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coren, S., & Girgus, J. S. (1972). Illusion decrement in intersecting line figures. *Psychonomic Science*, 26(2), 108–110, <https://doi.org/10.3758/BF03335451>.
- Coren, S., Girgus, J. S., Erlichman, H., & Hakstian, A. R. (1976). An empirical taxonomy of visual illusions. *Perception & Psychophysics*, 20(2), 129–137, <https://doi.org/10.3758/BF03199444>.

- Cretenoud, A. F., Francis, G., & Herzog, M. H. (2020). When illusions merge. *Journal of Vision*, 20(8), 1–15, <https://doi.org/10.1167/JOV.20.8.12>.
- Cretenoud, A. F., Grzeczowski, L., Bertamini, M., & Herzog, M. H. (2020). Individual differences in the Müller-Lyer and Ponzo illusions are stable across different contexts. *Journal of Vision*, 20(6), 1–14, <https://doi.org/10.1167/JOV.20.6.4>.
- Cretenoud, A. F., Karimpur, H., Grzeczowski, L., Francis, G., Hamburger, K., & Herzog, M. H. (2019). Factors underlying visual illusions are illusion-specific but not feature-specific. *Journal of Vision*, 19(14), 1–21, <https://doi.org/10.1167/19.14.12>.
- Day, R. H. (1962). The effects of repeated trials and prolonged fixation on error in the Müller-Lyer figure. *Psychological Monographs: General and Applied*, 76(14), 1–19, <https://doi.org/10.1037/h0093815>.
- Emery, K. J., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017a). Variations in normal color vision. VI. Factors underlying individual differences in hue scaling and their implications for models of color appearance. *Vision Research*, 141, 51–65, <https://doi.org/10.1016/j.visres.2016.12.006>.
- Emery, K. J., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017b). Variations in normal color vision. VII. Relationships between color naming and hue scaling. *Vision Research*, 141, 66–75, <https://doi.org/10.1016/j.visres.2016.12.007>.
- Festinger, L., White, C. W., & Allyn, M. R. (1968). Eye movements and decrement in the Müller-Lyer illusion. *Perception & Psychophysics*, 3(5), 376–382, <https://doi.org/10.1007/BF03396511>.
- Fisher, R. A. (1992). Statistical Methods for Research Workers. In *Breakthroughs in Statistics*. (pp. 66–70). New York, NY: Springer, https://doi.org/10.1007/978-1-4612-4380-9_6.
- Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78, <https://doi.org/10.1016/j.paid.2016.06.069>.
- Grzeczowski, L., Clarke, A. M., Francis, G., Mast, F. W., & Herzog, M. H. (2017). About individual differences in vision. *Vision Research*, 141, 282–292, <https://doi.org/10.1016/j.visres.2016.10.006>.
- Grzeczowski, L., Roinishvili, M., Chkonia, E., Brand, A., Mast, F. W., Herzog, M. H., . . . Shaqiri, A. (2018). Is the perception of illusions abnormal in schizophrenia? *Psychiatry Research*, 270, 929–939, <https://doi.org/10.1016/j.psychres.2018.10.063>.
- Halpern, S. D., Andrews, T. J., & Purves, D. (1999). Interindividual variation in human visual performance. *Journal of Cognitive Neuroscience*, 11(5), 521–534, <https://doi.org/10.1162/089892999563580>.
- Happé, F. G. E. (1996). Studying weak central coherence at low levels: Children with autism do not succumb to visual illusions. A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 37(7), 873–877, <https://doi.org/10.1111/j.1469-7610.1996.tb01483.x>.
- Iglewicz, B., & Hoaglin, D. (1993). *Volume 16. How to detect and handle outliers*. Milwaukee, WI: ASQC Press.
- Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.
- Judd, C. H. (1902). Practice and its effects on the perception of illusions. *Psychological Review*, 9(1), 27–39, <https://doi.org/10.1037/h0073071>.
- Kiely, K., & Anstey, K. (2017). Common Cause Theory in Aging. In *Encyclopedia of Geropsychology* (pp. 559–569). Singapore: Springer Science + Business Media.
- King, D. J., Hodgekins, J., Chouinard, P. A., Chouinard, V. A., & Sperandio, I. (2017). A review of abnormalities in the perception of visual illusions in schizophrenia. *Psychonomic Bulletin and Review*, 24(3), 734–751, <https://doi.org/10.3758/s13423-016-1168-5>.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163, <https://doi.org/10.1016/j.jcm.2016.02.012>.
- Leibowitz, H. W., Brislin, R., Perlmutter, L., & Hennessy, R. (1969). Ponzo perspective illusion as a manifestation of space perception. *Science*, 166(3909), 1174–1176, <https://doi.org/10.1126/science.166.3909.1174>.
- Liljequist, D., Elfving, B., & Skavberg Roaldsen, K. (2019). Intraclass correlation – A discussion and demonstration of basic features. *Plos One*, 14(7), 1–35, <https://doi.org/10.1371/journal.pone.0219854>.
- Lindenberger, U., & Ghisletta, P. (2009). Cognitive and Sensory Declines in Old Age: Gauging the Evidence for a Common Cause. *Psychology and Aging*, 24(1), 1–16, <https://doi.org/10.1037/a0014986>.
- Manning, C., Morgan, M. J., Allen, C. T. W., & Pellicano, E. (2017). Susceptibility to Ebbinghaus and Müller-Lyer illusions in autistic children: a comparison of three different methods. *Molecular Autism*, 8, 1–18, <https://doi.org/10.1186/s13229-017-0127-y>.
- Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is

- good experimental practice? *Vision Research*, 141, 4–15, <https://doi.org/10.1016/j.visres.2017.11.001>.
- Ninio, J. (2014). Geometrical illusions are not always where you think they are: a review of some classical and less classical illusions, and ways to describe them. *Frontiers in Human Neuroscience*, 8:856, 1–21, <https://doi.org/10.3389/fnhum.2014.00856>.
- Partos, T. R., Cropper, S. J., & Rawlings, D. (2016). You Don't See What I See: Individual Differences in the Perception of Meaning from Visual Stimuli. *Plos One*, 11(3), 1–26, <https://doi.org/10.1371/journal.pone.0150615>.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Peterzell, D. H. (2016). Discovering Sensory Processes Using Individual Differences: A Review and Factor Analytic Manifesto. *Electronic Imaging*, 2016(16), 1–11, <https://doi.org/10.2352/ISSN.2470-1173.2016.16HVEI-112>.
- Peterzell, D. H. (2020). Individual Differences in Perceptual Organization: Reanalyzing Thurstone's classic (1944) data and rediscovering factors for geometrical illusions, perceptual switching, and holistic 'Gestalt' closure. *Journal of Vision*, 20(11), 140, <https://doi.org/10.1167/jov.20.11.140>.
- Peterzell, D. H., Scheffrin, B. E., Tregear, S. J., & Werner, J. S. (2000). Spatial frequency tuned covariance channels underlying scotopic contrast sensitivity. In *Vision Science and its Applications: OSA Technical Digest*. Washington, D.C.: Optical Society of America.
- Peterzell, D. H., Serrano-Pedraza, I., Widdall, M., & Read, J. C. A. (2017). Thresholds for sine-wave corrugations defined by binocular disparity in random dot stereograms: Factor analysis of individual differences reveals two stereoscopic mechanisms tuned for spatial frequency. *Vision Research*, 141, 127–135, <https://doi.org/10.1016/j.visres.2017.11.002>.
- Piaget, J. (1961). *Les Mécanismes perceptifs, Modèles probabilistes, Analyse génétique, Relations avec l'intelligence* (2nd ed.). Paris: Presses universitaires de France.
- Predebon, J. (2006). Decrement of the Müller-Lyer and Poggendorff illusions: The effects of inspection and practice. *Psychological Research*, 70(5), 384–394, <https://doi.org/10.1007/s00426-005-0229-6>.
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, <https://www.r-project.org/>.
- Robinson, J. O. (1968). Retinal inhibition in visual distortion. *British Journal of Psychology*, 59(1), 29–36, <https://doi.org/10.1111/j.2044-8295.1968.tb01113.x>.
- Roff, M. (1953). A Factorial Study of Tests in the Perceptual Area. *Psychometric Monographs*, 8, 1–41.
- Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience*, 14(1), 28–30, <https://doi.org/10.1038/nn.2706>.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428, <https://doi.org/10.1037/0033-2909.86.2.420>.
- Song, C., Schwarzkopf, D. S., & Rees, G. (2011). Interocular induction of illusory size perception. *BMC Neuroscience*, 12(1), 1–9, <https://doi.org/10.1186/1471-2202-12-27>.
- Taylor, T. R. (1974). A factor analysis of 21 illusions: The implications for theory. *Psychologia Africana*, 15, 137–148.
- Taylor, T. R. (1976). The factor structure of geometric illusions: A second study. *Psychologia Africana*, 16, 177–200.
- Thurstone, L. L. (1944). *A factorial study of perception*. Chicago, IL: The University of Chicago Press.
- Todorović, D., & Jovanović, L. (2018). Is the Ebbinghaus illusion a size contrast illusion? *Acta Psychologica*, 185, 180–187, <https://doi.org/10.1016/j.actpsy.2018.02.011>.
- Tulver, K. (2019). The factorial structure of individual differences in visual perception. *Consciousness and Cognition*, 73, 1–8, <https://doi.org/10.1016/j.concog.2019.102762>.
- Tulver, K., Aru, J., Rutiku, R., & Bachmann, T. (2019). Individual differences in the effects of priors on perception: A multi-paradigm approach. *Cognition*, 187(3), 167–177, <https://doi.org/10.1016/j.cognition.2019.03.008>.
- Wagner, D. A. (1977). Ontogeny of the Ponzo illusion: Effects of age, schooling, and environment. *International Journal of Psychology*, 12(3), 161–176, <https://doi.org/10.1080/00207597708247386>.
- Webster, M. A., & MacLeod, D. I. A. (1988). Factors underlying individual differences in the color matches of normal observers. *Journal of the Optical Society of America A*, 5(10), 1722–1735, <https://doi.org/10.1364/JOSAA.5.001722>.
- Wexler, M. (2018). Multidimensional internal dynamics underlying the perception of motion. *Journal of Vision*, 18(5), 1–12, <https://doi.org/10.1167/18.5.7>.

Wexler, M., Duyck, M., & Mamassian, P. (2015). Persistent states in vision break universality and time invariance. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(48), 14990–14995, <https://doi.org/10.1073/pnas.1508847112>.

Yildiz, G. Y., Sperandio, I., Kettle, C., & Chouinard, P. A. (2021). Interocular transfer effects of linear perspective cues and texture gradients in the perceptual rescaling of size. *Vision Research*, *179*, 19–33, <https://doi.org/10.1016/j.visres.2020.11.005>.