Individual differences distinguish mechanisms underlying contextual size illusions

Dietrich S. Schwarzkopf, Xinran A. Yu, & Livia F. Fischer

School of Optometry & Vision Science, University of Auckland, New Zealand

Abstract

Previous research suggests the magnitudes of the Ebbinghaus, Delboeuf, Ponzo, and tilt illusion all depend on the cortical distance between the neural representations of target stimuli and the surrounding context. However, several psychophysical studies found no compelling association between these illusions, calling this hypothesis into question. Here we ask if these discrepant reports could arise from methodological differences between these studies. We ran a battery of visual size illusion and basic discrimination tasks with carefully matched geometric properties, using a classical forced choice design. Ebbinghaus and Delboeuf illusion magnitudes are strongly correlated, consistent with the idea that they reflect the same underlying mechanism. Ponzo illusion magnitude also correlates with these two illusions, despite to a lesser extent, indicative that both shared and independent factors govern these illusions. Interestingly, the classical arrowhead version of the Mueller-Lyer illusion does not correlate with any of the other illusions or the ability to discriminate line length. This suggests that an altogether separate process underlies this perceptual effect. We further demonstrate that presenting stimuli briefly with central fixation critically affects measurements of the Ebbinghaus illusion. Taken together, our findings therefore highlight the importance of experimental parameters when probing relationships between perceptual effects and their links to neural processing.

Keywords

Spatial vision, perception, illusions, cortical distance, primary visual cortex, psychophysics, experimental design, correlations

Introduction

The human visual system is not a digital camera. Our interpretation of visual space is subject to many contextual effects that can manifest as perceptual illusions. These effects can arise from basic limitations of the neural architecture, such as the spatial resolution in peripheral vision is limited by having fewer neuronal resources than in central vision. In contrast, other illusions reflect the tasks the visual system has evolved to solve. For instance, navigating and interacting with a three-dimensional world involves interpreting the size of objects and their distance from the observer. Faced with a two-dimensional image that deliberately violates real-world relationships can lead to perceptual distortions.

Mechanisms mediating contextual illusions

Several illusions of spatial context have been linked to the functional architecture and/or response properties of the visual cortex. Starting with a seminal study (Murray et al., 2006), numerous reports have established that activity in human primary visual cortex (V1) reflects the apparent, not the veridical, eccentricity of stimuli (Fang et al., 2008; He et al., 2015; Ni et al., 2014), the apparent size of retinal afterimages (Sperandio et al., 2012), the repulsive perceptual effect of adapting to different stimulus sizes (Pooresmaeili et al., 2013), and the perceived separation in a dot variant of the Mueller-Lyer illusion (Ho & Schwarzkopf, 2022). One study using also suggested that effective connectivity between cortical representations of target and context stimuli correlates with the magnitude of the tilt illusion (Song, Schwarzkopf, Lutti, et al., 2013). Tackling this problem from a different perspective, we found that individual differences in the surface area of V1 correlate with the magnitude of the Ebbinghaus and Ponzo illusions (Schwarzkopf et al., 2011; Schwarzkopf & Rees, 2013), the tilt illusion (Song, Schwarzkopf, & Rees, 2013), and possibly the Delboeuf illusion (Moutsiana et al., 2016). These findings allude to a role of cortical magnification in mediating these contextual illusions: the magnitude and direction of the Ebbinghaus, Delboeuf, and tilt illusions are well explained by the cortical distance between target and context stimuli in V1 (Mareschal et al., 2010; Schwarzkopf & Rees, 2013; Urale & Schwarzkopf, 2023). One indication that the Ebbinghaus illusion is mediated by early stages of visual processing is that it is substantially reduced when targets and inducers are shown to different eyes (Song et al., 2011) and partly persists when inducers are masked from awareness by continuous flash suppression (Chen et al., 2018). This suggests a role for monocular channels.

The Ebbinghaus is often described as a “size contrast” effect. This is based on naïve interpretation of the effect in the classical configuration of this illusion: a circle surrounded by a context of small circles (inducers), appears larger than the same circle surrounded by large inducers (Figure 1A). However, this interpretation misses the fact that when testing each inducers configuration separately against a no-context control, the illusion is in fact an underestimation of target size for most stimulus parameters (Roberts et al., 2005; Schwarzkopf & Rees, 2013; Urale & Schwarzkopf, 2023). Rather than size contrast, the illusion is in fact a geometric effect of the inducers. The parameters exerting the most profound

influence on the illusion are the distance between the target and the inducers, and how much of the surround is covered by inducers (Roberts et al., 2005; Urale & Schwarzkopf, 2023). This has led to the suggestion that the Delboeuf illusion is in fact a special case of the Ebbinghaus: an Ebbinghaus illusion with an annulus of very small, densely packed inducers is effectively a Delboeuf stimulus, yielding a nigh identical profile of illusion magnitudes (Roberts et al., 2005). According to this interpretation, results for these two illusions should be very similar – that is, strongly correlated – across observers. The Ebbinghaus illusion is however more complex; the surrounding context comprises multiple stimuli with different sizes, which are likely to interact with each other as well as the target. Moreover, there have been various suggestions of purportedly higher-level modulators of the Ebbinghaus illusion, such as the similarity between targets and inducers (Choplin & Medin, 1999; Deni & Brigner, 1997). While some of these effects are conceivably due to low-level feature interactions, at least some studies suggest a role for semantic similarity (Coren & Enns, 1993). If true, this implicates factors beyond local geometric interactions mediated by V1 circuits. This would predict a substantial proportion of the variance in the Ebbinghaus illusion is unrelated to that of the Delboeuf illusion.

Similar questions arise about the Ponzo illusion (Figure 1C). Authors often consider the Ponzo to reflect the interpretation of distance in depth from flat two-dimensional images, a failure of size constancy. Such a process conceivably involves higher-order processing well beyond primary visual cortex, because it requires interpreting three-dimensional scenes and neurons with larger receptive fields than V1. This hypothesis is borne out by findings that the Ponzo illusion exhibits complete interocular transfer (Song et al., 2011) and that it is wiped out when rendering the context invisible by continuous flash suppression (Chen et al., 2018). We and others also showed that both the classical Ponzo, a corridor-type illusion with real-world scenes, and various steps in between, all show an inversion effect whereby the illusion is substantially stronger if the context implies the ground rather than the ceiling (Altan et al., 2023; Brislin, 1974; Poom, 2020). This finding is better explained by a higher-level interpretation of the images than a lower-level geometric interaction. While we found that the Ponzo illusion correlates with V1 surface area, just like the Ebbinghaus, this correlation was probably driven by an influential outlier and is therefore not robust (Schwarzkopf et al., 2011). Curiously, we also found no correlation between the Ebbinghaus and Ponzo illusion magnitudes themselves. Nevertheless, geometric parameters could theoretically play a role in the Ponzo. In the classical Ponzo configuration, the target which is typically perceived as smaller (and nearer to the observer) is generally farther from the contextual stimuli; this conforms with the effect of target-inducer distance in the Ebbinghaus and Delboeuf illusions (Roberts et al., 2005; Urale & Schwarzkopf, 2023). Even in real-world corridor-type versions of the Ponzo, the distance between the “near/smaller” target and landmarks in the contextual image is greater. Interestingly, bandpass filtering the contextual images to different spatial frequency ranges does not impact the illusion magnitude (Yildiz et al., 2021). This suggests that while a purely low-level mechanism for this illusion is unlikely, it does not rule out a contribution of such processes.

Finally, in the classical Mueller-Lyer illusion the lines are perceptually lengthened or shortened depending on the direction of arrowheads at the line ends (Figure 1D). This effect is often thought to involve higher-order analysis of three-dimensional scenes, such as interpreting the two configurations as edges/corners near or far from the observer (Gregory, 1966). Speaking against this are the facts that the Brentano version of the illusion manifests as a perceived position shift and correspondingly the mid-point of the line, which is at odds with interpreting the lines as a cue for depth. The illusion is also effective for simple dot variants devoid of any lines (Ho & Schwarzkopf, 2022). The perception of this dot variant is reflected in V1 responses (Ho & Schwarzkopf, 2022), and the illusion is reduced when adapting to low, but not high, spatial frequency patterns (Carrasco et al., 1986). Nevertheless, the current literature cannot conclusively rule out a role of higher-order processing in this illusion, especially the classical line and arrowhead version which is usually stronger than stripped down versions like the dot variant (Coren, 1970). As with the Ponzo and the Ebbinghaus, it is likely that multiple separate mechanisms contribute to the Mueller-Lyer.

Given this diversity, it seems inconceivable there could be a common factor to all spatial contextual illusions, although this has not stopped researchers from searching for such factors (Axelrod et al., 2017; Cretenoud et al., 2021; Grzeczkowski et al., 2017; Jastrzębowska et al., 2023). A large psychophysical study on simultaneous contrast tasks found that the magnitudes of these illusions are mostly uncorrelated between stimulus dimensions (Bosten & Mollon, 2010), something in line with our own findings comparing size and luminance perception (Song, Schwarzkopf, & Rees, 2013). Often only versions of the same illusion (e.g., different configurations of the Ebbinghaus) share any significant variance at all (Grzeczkowski et al., 2017). However, many of these reports used very distinct stimulus parameters, making no effort to match conditions. Such an experimental design is reasonable for asking if there are any correlations between illusions in the broadest sense. For effects that arise in higher decision-making stages of cognition the exact stimulus parameters may be largely irrelevant. However, when testing hypotheses about the involvement of circuits in the early visual system, this is fundamentally flawed. If both the Delboeuf and Ebbinghaus illusion depend on the cortical distance between target and inducers, it is imperative to ensure that any comparison between them controls extraneous factors rather than presenting each illusion at different visual field locations with different geometric parameters.

Experiments investigating these visual cortical processes also require brief stimulus presentation and stable eye fixation. Both are essential for ensuring that stimuli are presented in fixed locations of the visual field and thus target the appropriate neuronal populations in retinotopic visual cortex. Yet, most studies exploring the relationship between different illusions used unconstrained experimental designs – again, a reasonable approach for studying the higher-order cognitive aspects of these tasks, but highly problematic for studying perceptual processing.

Measuring illusion magnitude

The ways how illusion magnitudes are measured is also critical. Any behavioural measures, including those quantifying perceptual experience, reflect the output of a whole brain working in concert. This ranges from multiple steps of image processing, making perceptual inferences, and eventually producing a motor output to report the judgement. Therefore, no illusion is plausibly mediated solely by a single process. Numerous paradigms exist for quantifying the magnitude of an observer's perceptual illusions, each of which involves cognitive confounds to some degree: this includes adjustment/reproduction tasks (e.g. (Ho & Schwarzkopf, 2022; Jastrzębowska et al., 2023)), forced-choice procedures (e.g. (Schwarzkopf & Rees, 2013)), confidence ratings (Gallagher et al., 2019), and drift diffusion models (Sánchez-Fuenzalida et al., 2023). Adjustment task may be particularly prone to demand effects because they rely on a subjective criterion for when there is a perceptual match between a test and reference stimulus. However, it is now widely recognised that even classical forced-choice procedures – unbiased for measuring objective performance like acuity or detection thresholds – are susceptible to decision-making confounds. This has led to the development of more complex procedures based on comparing the difference between differences to disentangle perceptual from cognitive biases (Finlayson et al., 2017; Gallagher et al., 2019; Jogan & Stocker, 2014; M. Morgan et al., 2012; M. J. Morgan et al., 2013; Patten & Clifford, 2015).

Nevertheless, the classical two-alternative forced choice (2AFC) design in which observers must make a perceptual judgement between a constant reference and a variable test stimulus remain widely used and efficient methods for quantifying illusion magnitudes. Provided the observer's expectations or preconceived notions of what the measurement reflects (such as suggestions that an illusion might be stronger/weaker in certain demographics) can be minimised, the 2AFC design is arguably still the most optimal method. One way to achieve this is by randomly interleaving different conditions, randomising when/where reference and test stimuli are presented, and using a speeded design to prevent the observer from relying on non-perceptual cues in their decision.

Importantly, most of the studies failing to find correlations between most contextual illusions used adjustment procedures (Cretenoud et al., 2021; Cretenoud, Francis, et al., 2020; Cretenoud, Grzeczkowski, et al., 2020; Grzeczkowski et al., 2017; Jastrzębowska et al., 2023). Some of these experiments were conducted online or outside lab environments, allowing minimal control over stimulus characteristics and eye gaze. Generally, adjustment tasks can be done in considerably less time than forced-choice paradigms. However, this confounds any conclusions we can draw about these findings. One study compared the results of adjustment and a 2AFC design and argued they were strongly correlated (Cretenoud et al., 2021). The correlations ranged between 0.6-0.67, meaning that between half to two thirds of the variance remained unexplained. It is conceivable that the correlation in these experiments largely reflects the higher-level decision processes but that the residuals – where the two measurements diverge – are due to the underlying neural perceptual processing. While this is

not explicitly stated in the methods[1], it is likely that these experiments did not control eye gaze and used unlimited stimulus presentations in the 2AFC tasks. As explained, this would further confound any conclusions we can draw from these experiments about early visual processing.

The present study

Here we set out to quantify the individual differences in these contextual size modulations, specifically the Ebbinghaus, Delboeuf, Ponzo, and Mueller-Lyer illusions. We tested a battery comprising these size illusions, using a conventional two-alternative forced choice (2AFC) procedure with brief stimulus presentations and fixation instructions. We also included two objective discrimination tasks for judging the width of circles and the length of lines without any illusion-inducing contexts. This is because our previous research indicated associations between the objective acuities and subjective illusion magnitudes and associated spatial selectivity of visual neuronal populations (Moutsiana et al., 2016; Song et al., 2015; Song, Schwarzkopf, & Rees, 2013). The stimulus conditions in all these tasks were closely matched. For example, target stimuli were identical filled circles for testing the Ebbinghaus, Delboeuf, and Ponzo illusions as well as the corresponding circle width discrimination tasks, and the length of the lines in the Mueller-Lyer illusion and the corresponding line length discrimination task matched the diameters of the circles in the other tasks (Figure 1A-F). In the same vein, the distances between targets and inducers were comparable between the different illusions. This design therefore maximised the chance of finding associations between these perceptual effects; the influence of any extraneous stimulus differences should be minimised.

We tested a series of preregistered hypotheses aimed to better understand the mechanisms mediating these illusions. Critically, under the proposition that the Ebbinghaus and Delboeuf illusions are effectively the same phenomenon (Roberts et al., 2005), we predicted that these two effects are strongly correlated, directly contradicting other recent reports (Jastrzębowska et al., 2023).

Moreover, in a second experiment, we further assessed the effect of task and stimulus parameters. We compared Ebbinghaus illusion magnitudes between adjustment tasks, a 2AFC task with unlimited stimulus presentation and free eye movements, and a classical 2AFC task with brief stimulus presentation and fixation instructions, as used in our earlier studies investigating brain-perception links for this illusion (Schwarzkopf & Rees, 2013). We also compared the effect of using two different stimulus parameters because the stimulus details described in our earlier study contained an error[2], meaning researchers seeking to replicate those experiments following the methods would inadvertently use different stimuli than our original study (Jastrzębowska et al., 2023).

---

[1] We contacted the authors of that study about this point but have not received a response to our inquiry.

[2] Corrections for this typo have been posted as comments on the original article and on PubPeer. A formal correction was rejected by the journal because they consider this error "minor".

Materials and Methods

Procedures were preregistered on the Open Science Framework for both Experiment 1 (https://osf.io/bhgud) and Experiment 2 (https://osf.io/5ey8h).

Participants

Thirteen observers (ages 18-45, 10 female, 3 left-handed) were recruited for this study. Three of these were the authors, but the remaining observers were naïve to the purpose of the experiments. One observer was excluded from the whole study based on preregistered criteria (see below). Another observer was excluded from Experiment 2 only. Observers gave written informed consent, and all procedures were approved by the University of Auckland Human Participants Research Ethics Committee. Observers received a NZ$25 supermarket voucher for their participation in both experiments.

Experiment 1

Stimuli

Stimuli were closely matched between the different experimental conditions. In all tests, the fixation dot was a white disc with diameter of 0.15° visual angle, presented in the centre of the screen. All other stimuli were a dark grey (8-bit intensity 64). The line width of any unfilled circles and line stimuli was 0.1°. All target stimuli were centred at an eccentricity of 4.65° to the left and right of fixation. The reference stimulus size (circle diameter or line length) was always 2.06°. The width or length of test stimuli varied on a binary logarithmic scale. The reference and test stimuli could appear either on the left or right of fixation. The main differences between the experimental conditions were in the design of the contextual stimuli (Figure 1A-F):

1. *Circle width discrimination:* Only the target circles were presented. The stimulus design was identical to the control condition in our earlier study (Schwarzkopf & Rees, 2013), but note that the methods in that article incorrectly described the radius of the circle stimuli as diameters.
2. *Ebbinghaus illusion:* The target stimuli were filled circles, surrounded by other filled circles, the inducers, that were placed in equal angular steps, starting from the 3 o'clock position. Sixteen small inducers (0.52° diameter) were centred 1.86° from the centre of the reference stimulus. Six large inducers (4.14° diameter) were centred 4.34° from the centre of the test stimulus. These stimuli were also identical to the illusion stimuli in our earlier study (Schwarzkopf & Rees, 2013).
3. *Delboeuf illusion:* The target stimuli were filled circles, surrounded by rings. The ring around the reference and test stimuli had a radius of 1.6° and 4.34°, respectively.

4. *Ponzo illusion:* The target stimuli were filled circles. They were placed within a context of two converging lines that could either converge towards the left or right. The lines started and ended 5.9° horizontally from the vertical meridian in both directions. To converge, their ends were positioned above and below the horizontal meridian at a distance 4.3° and 1.24°, respectively. The reference and test stimuli were at the narrower and wider ends, respectively.
5. *Line length discrimination:* The target stimuli were vertical lines. They bisected horizontal lines (1.3° long) at each end.
6. *Mueller-Lyer illusion:* The target stimuli were vertical lines. At each end, they were joined to two 0.65° long lines angled at ±45° relative to vertical, forming an arrowhead. In the test and reference stimuli, the lines were directed towards or away from the horizontal meridian, respectively.

Procedure

After giving consent, participants were seated in front of a computer screen at a viewing distance of 57 cm, with their heads stabilised by a chin and forehead rest. The general experimental design was a 2AFC task with an identical trial structure for all illusion and discrimination conditions. We matched stimulus parameters between the conditions as much as reasonably feasible. The task involved making a perceptual judgement about which target circle was larger or which line was longer, respectively. There were always two stimuli presented simultaneously within each trial, with one being a constant reference (diameter/length of circles/lines: 2.06°) and the other a test stimulus that varied across trials.

The experimental conditions were presented in the following sequence: whether a participant started with the line judgements (line length discrimination and Mueller-Lyer illusion) or the circle judgements (circle width discrimination, and Ebbinghaus, Delboeuf, and Ponzo illusions) was pseudo-randomised. However, the objective discrimination tasks for these judgements always came first before testing the corresponding illusions. The order in which the three circle illusions were tested was counterbalanced across participants, resulting in 6 possible permutations.

For each condition, the participant first performed a few demo trials in which the stimuli were presented until the participant's keyboard response. This allowed the experimenter to explain the task. This was followed by more practice trials in which the trial sequence was as in the actual experiment, with 100 ms stimulus presentations. Once the participant confirmed they understood the task, the actual experiment began.

Trials stared with a brief 500 ms fixation period where only the fixation dot was presented on a black screen. This was followed by a 100 ms stimulus presentation, after which the screen turned black. The participant must then make a response to judge whether the left or right stimulus was larger/longer by pressing either the left or right arrow key. The next trial then started immediately.

The 160 trials were divided into four blocks of 40 trials each. There were four trial types: the two interleaved staircases (starting above or below true equality) and the reference stimulus could either appear to the left or the right of fixation. These four trial types were repeated 10 times per block in a pseudo-randomised order. Before each block, there was a rest screen reminding the participant of the task instructions and informing them of the number of blocks they had already completed. They commenced the block by pressing any key on the keyboard.

Stimulus levels were expressed in binary logarithms of the ratio between the test and reference size (diameter for circles, length for lines). This linearises stimulus values to account for Weber's Law and allows the use of parametric statistical tests, which would not be applicable in the case of raw size ratios.

Test stimulus levels were controlled by two interleaved 1-down, 1-up staircase procedures that converged onto the point of subjective equality (PSE). One staircase started at a higher stimulus level (0.4 log units, wider circle or longer line) than true equality (0 log units), and the other staircase started at a lower level (-0.4 log units, narrower circle or shorter line). The staircase increased or decreased the stimulus level following the response in each trial. Initially, the increment/decrement was at a larger step size of 0.1 log units. Once five reversals had occurred in a given staircase, the step size was reduced to 0.05 log units. Moreover, from this point on, there was also a 10% probability on each trial that the stimulus level was increased or decreased (depending on the starting point of the given staircase) by 0.15 log units. These "jitter trials" prevented the stimulus level from continuously hovering around the PSE, as this could result in disengagement from the task.

Data analysis

Observers completed 160 trials per experimental condition. Data were then analysed by calculating the proportion of trials in which the test stimulus was chosen for each stimulus level used (collapsing across both staircases in each condition and irrespective of which side the test was presented). We then fit a cumulative Gaussian function to these data points using the summed squared residuals approach, taking into account the robustness of each data point by weighting them by the number of trials at that stimulus level. We inferred the PSE from the stimulus level where the curve crossed the 50% response level (centroid of the psychometric curve). We also estimated discrimination uncertainty (conceptually a threshold or just noticeable difference) by determining the bandwidth of the psychometric curve.

Our preregistered plan was to exclude observers who failed to complete all experimental conditions; however, all observers completed all conditions. Further, we excluded participants for whom the bandwidth (technically, the σ parameter) for any psychometric curve fit exceeded 0.78 log units. Beyond this bandwidth, assuming the PSE is zero an observer would only score 10% or 90%, respectively, for stimulus levels of +/- 1 log units (i.e., test stimuli half or twice as large as the reference). We reasoned that this constitutes an implausibly poor performance for these simple visual discrimination tasks in observers with supposedly normal, healthy vision. All excluded participants were to be replaced to ensure correct

counterbalancing of conditions (see below for deviation from these criteria). One observer was excluded based on this criterion.

The PSE (centroids of psychometric curves) for the illusion conditions and the discrimination abilities (bandwidths of psychometric curves) for the two objective discrimination tasks (circle width and line length) were compared using Pearson's r correlation. We tested all hypotheses using directional (one-tailed) tests, using a Bayes Factor with a default prior as implemented in JASP (0.17.2.1).

Sampling plan

Our minimum sample size was N=6 because that is required to include all sequences for circle judgement illusions. We then incremented the sample size by collecting batches of 6 and testing the stopping criterion. This criterion was based on hypothesis I_1, that is, whether the one-tailed Pearson's correlation between the Ebbinghaus and Delboeuf illusion magnitudes yielded a Bayes Factor of 10 or 0.1, with data either supporting the alternative or null hypothesis, respectively. We chose this hypothesis for the stopping criterion because we expected these two illusions to reflect the same underlying neural process, yet previous studies failed to find any strong correlation between them (Jastrzębowska et al., 2023). Our maximum sample size in case the stopping criterion was not reached was N=90. However, the criterion was reached after two batches were collected, that is, at N=12.

Preregistered hypotheses

We tested the following hypotheses of between-subject correlations between psychometric parameters estimated for individual participants:

Hypothesis I_1: Ebbinghaus and Delboeuf illusion magnitudes (PSEs) correlate positively, because they are effectively the same process.

Hypothesis I_2: Ponzo illusion magnitude does not correlate (or only weakly) with magnitudes of the other illusions, because it reflects higher-level processing of three-dimensional scenes.[3]

Hypothesis S_1: Ebbinghaus and Delboeuf illusion magnitudes each correlate positively with bandwidth for objective circle width discrimination.

Hypothesis S_2: Ponzo illusion magnitude does not correlate positively (or only weakly) with bandwidths for objective circle width discrimination.

Hypothesis S_3: Bandwidths for the objective discrimination tasks (circle width and line length) correlate positively because they both rely on local V1 processing (such as receptive field size).

[3] Note that this hypothesis contradicts exploratory hypothesis M_3 below as we had no specific *a priori* predictions for the Mueller-Lyer illusion. If M3 had been confirmed, we would have expected I_2 to be only partially confirmed.

Hypothesis M_1: Mueller-Lyer illusion magnitude correlates positively with bandwidth for line length discrimination.

Exploratory Hypothesis M_2: Mueller-Lyer illusion magnitude correlates positively with the Ebbinghaus and Delboeuf illusion magnitudes if they share similar early neural mechanisms. Some previous findings suggest that this illusion could be mediated by local interactions in V1 (Carrasco et al., 1986; Ho & Schwarzkopf, 2022; M. J. Morgan & Glennerster, 1991), but this has not been tested using classical arrow-type Mueller-Lyer stimuli. Conversely, the correlation is only weak if this illusion reflects higher-order scene processing.

Exploratory Hypothesis M_3: Mueller-Lyer illusion magnitude correlates positively with the Ponzo illusion, if they both reflect the same process of interpreting 3D scenes.

Experiment 2

Stimuli

One of the Ebbinghaus stimulus designs (*Grey&Large*) was identical to Experiment 1 (and a direct replication of our earlier study (Schwarzkopf & Rees, 2013)). The other stimulus design (*White&Small*) was based on a replication study (Jastrzębowska et al., 2023), which was based on the incorrect methods in our original study (although we note that correct example stimuli were included in that study). Specifically, the spatial separation and eccentricities were identical to the other condition, but the filled circle stimuli were half as wide. This replication study also presented stimuli in white instead of dark grey. For Grey&Large circle stimuli, the fixation dot was a white dot as described in the preceding experiment, but for White&Small circle stimuli the fixation dot was black, because it fell inside an inducer.

Procedure

There were no demo or practice trials in this experiment, because participants had already done a very similar Ebbinghaus measurement in Experiment 1 immediately prior to this experiment. Instead, participants were shown the two possible stimulus parameters (Figure 2A-B) as an image on the screen, and the experimenter explained again what the perceptual judgement entailed.

Participants then performed three tasks, with the order of tasks counterbalanced across participants, resulting in six permutations:

1. *Fast-2AFC:* This was essentially identical to the Ebbinghaus measurement in the preceding experiment, except for the inclusion of the White&Small circle condition. Participants were explicitly instructed to fixate on the white dot in the centre of the screen. There were 320 trials, divided into blocks of 80 (160 trials per stimulus condition, i.e. the same number as in the preceding experiment where there was only one condition).
2. *Unconstrained-2AFC:* This was the same as the Fast-2AFC task, but there was no fixation dot shown and the illusion stimuli remained on screen until the participant

responded with the keyboard. They were explicitly told that they are free to move their eyes and inspect the stimuli as they please.

3. *Adjustment:* Stimuli appeared on the screen without a fixation dot. Participants then used the up and down arrow key on the keyboard to change the width of the test circle (surrounded by large inducers) until it was a perceptual match for the width of the reference circle. Once they were satisfied with the match, they pressed the space bar to move to the next trial. There were 20 trials in this test, 10 for each of the stimulus conditions. The side where the reference and test stimuli were presented was pseudo-randomised in each trial.

Stimulus values in the two 2AFC tasks were controlled by interleaved staircase procedures, just as in Experiment 1, except that there were independent staircases for the two stimulus conditions. In the Adjustment task, the initial radius of the test stimulus in each trial was drawn from a Gaussian distribution with a standard deviation of 0.4 log units.

Data analysis

Data from the two 2AFC tasks were analysed as in Experiment 1. For the Adjustment task, the illusion magnitude in this experiment for each stimulus condition was calculated as the arithmetic mean across the final adjusted size of the test stimulus from all trials. The paradigm also collected information about the time course of the Adjustment task in each trial, but this is irrelevant for the present study and will not be discussed further.

Mean illusion magnitudes were compared using 2x2 repeated-measures analysis of variance using stimulus condition and task as factors. Post-hoc contrasts and paired two-tailed t-tests compared individual conditions/tasks. The statistical evidence was based on default Bayes Factors. Correlations were calculated based on Pearson's r, evaluating statistical evidence based on default Bayes Factors. All these analyses were conducted in JASP 0.17.2.1 (JASP Team, 2023), using default priors.

The magnitude of correlations between the Adjustment and 2AFC tasks was compared using the Williams/Steiger equation for comparing two correlations with one shared variable (Howell, 1997). Statistical inference was based on the 95% confidence limit derived by the bootstrap percentile with 10,000 iterations. A significant difference between correlations is when the 95%CI does not overlap zero. To compare correlations between stimulus conditions for the separate tasks, we bootstrapped the difference in correlations using 10,000 iterations and calculated the 95% confidence level.

Because we reused the same participants as in Experiment 1, the sampling plan was determined by that experiment. We also applied the same exclusion criteria for Experiment 2, that is, if the bandwidth (technically, σ) of the psychometric curve fit for any of the experimental conditions exceeded 0.78 log units they would be excluded (without replacement, because recruitment was via Experiment 1). Based on this criterion, one observer was excluded, leaving only N=11 for Experiment 2.

Preregistered hypotheses

Hypothesis H_1: Illusion magnitude differs between the Adjustment task and the 2AFC tasks, specifically:

- *H_{1A}:* Illusion magnitude differs between Fast-2AFC and Adjustment.
- *H_{1B}:* Illusion magnitude differs between Fast-2AFC and the Unconstrained-2AFC.

Hypothesis H_2: Illusion magnitudes correlate between tasks as follows:

- *H_{2A}:* Illusion magnitudes correlate between Adjustment and Unconstrained-2AFC.
- *H_{2B}:* Illusion magnitudes correlate less strongly between Adjustment and Fast-2AFC.

Hypothesis H_3: Overall, illusion magnitude differs between the stimulus parameters irrespective of task. Qualitatively similar effects are found comparing stimulus parameters in each task separately.

Deviations from preregistration

Our preregistered protocol also included a fourth hypothesis about the correlations between stimulus parameters. However, after data collection had commenced we noticed that one sub-hypothesis (H_{4B}) in the preregistration was phrased incorrectly, effectively repeating hypothesis H_{2B}. The methods also did not contain the necessary description for the statistical comparison of these independent correlations. Even though this was intended to be a preplanned analysis, we therefore decided to label this as exploratory to avoid the impression that the preregistration was altered after the results were known.

We also adjusted our exclusion criteria slightly. It turns out that the Fast-2AFC task with the White&Small stimulus condition was very difficult for most observers. Several reported this condition as only a momentary flash, making it hard to make a perceptual judgement. One observer (P5) exceeded the exclusion criterion of bandwidth $\sigma>0.78$ for this condition ($\sigma=0.85$). However, this larger bandwidth was consistent with the rest of the sample and their other results accorded well with the rest of the sample. We therefore decided to keep this observer in the study. Control analysis confirmed that their exclusion would not have substantially altered the conclusions.

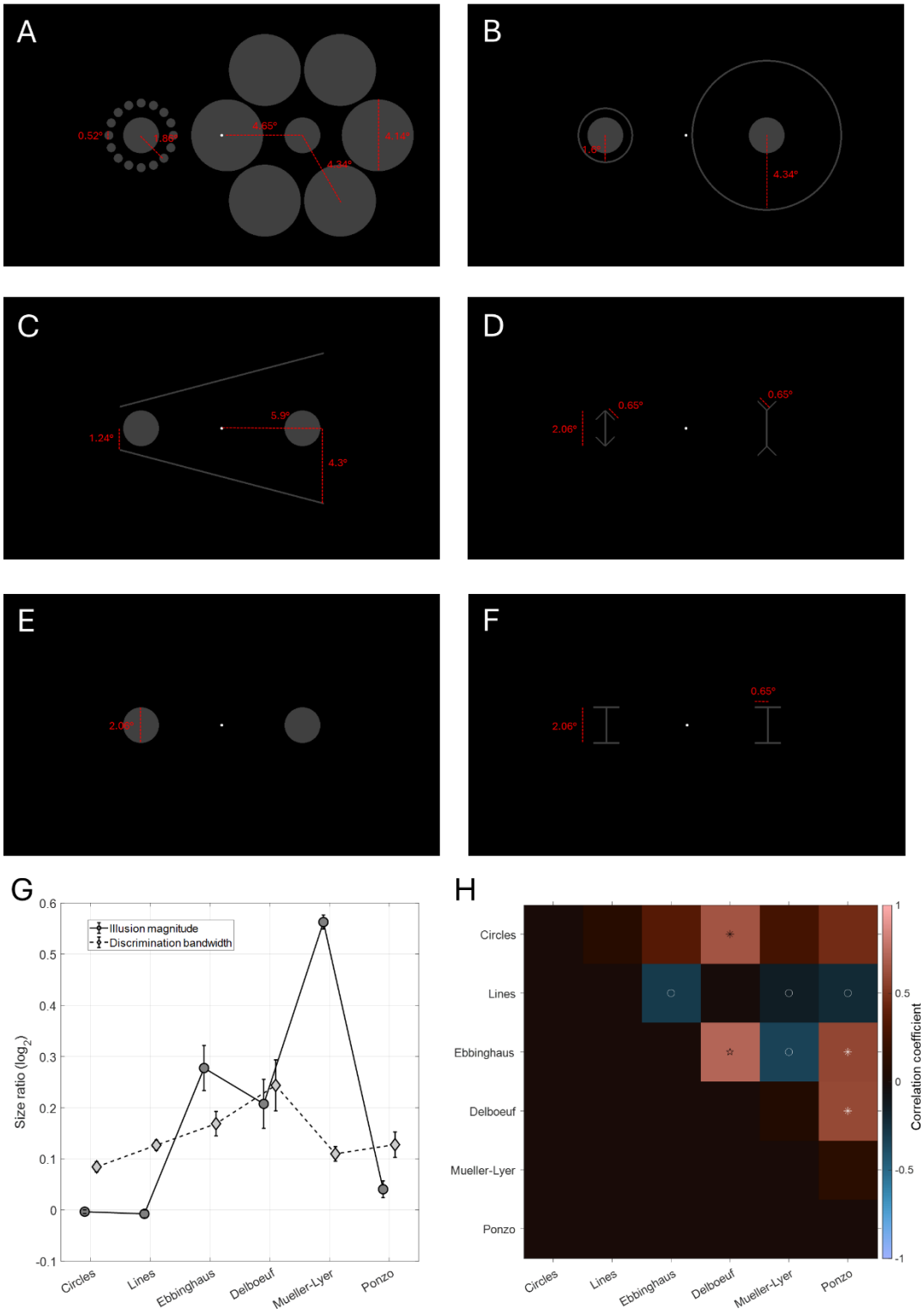

Figure 1. Stimuli (A-F) and results (G-H) of Experiment 1. In separate experimental runs, we estimated the illusion magnitude/perceptual bias and discrimination sensitivity for the Ebbinghaus (A), Delboeuf (B), Ponzo (C), Mueller-Lyer (D) illusions, or the discrimination of circle width (E) or line length (F). Red lines and numbers indicate dimensions in degrees of visual angle. These were not part of the presented stimuli. G. Average illusion magnitude (dark discs, solid lines) and discrimination bandwidths (light discs, dotted lines) for each condition. Error bars denote ±1 standard error of the mean across observers. H. Correlation matrix comparing the magnitudes for the four illusions and the discrimination bandwidths for judging circle width or line length, respectively. Colours indicate strength of correlation. In each cell, symbols denote Bayes Factors: asterisk: $BF_{+0}>3$, pentagram: $BF_{+0}>10$, circles $BF_{+0}<⅓$.

Results

Across two experiments, we aimed to better understand the discrepancy between previous studies in terms of the factors governing visual contextual size illusions. We quantified individual differences in illusion magnitudes and tested the effect of task and stimulus parameters on the Ebbinghaus illusion.

Illusions with circle targets are related

Our first experiment used a battery of tasks to compare the magnitudes of our range of illusions. Averaged across observers, the illusion magnitudes varied considerably, with by far the strongest effect for the Mueller-Lyer and the weakest for the Ponzo illusion (Figure 1G), with the Ebbinghaus and Delboeuf magnitude at similar intermediate levels. We also quantified the mean bandwidths of psychometric curves as a reflection of objective discrimination thresholds. This was largest (worst) for the Delboeuf and slightly elevated for the Ebbinghaus, but at similar levels for the other illusions and the objective discrimination tasks (Figure 1G).

Next, we calculated the pairwise correlations between the different illusions and the bandwidths for objective circle and line discrimination (Figure 1H). This revealed a strong correlation between the Ebbinghaus and Delboeuf illusion ($r=0.716$, $BF_{+0}=15.11$), confirming our primary hypothesis I_1. Surprisingly, the Ponzo illusion also correlated positively with both the Ebbinghaus and Delboeuf illusions (Ebbinghaus: $r=0.582$, $BF_{+0}=4.027$; Delboeuf: $r=0.597$, $BF_{+0}=4.533$), contradicting earlier findings by ourselves and others (Jastrzębowska et al., 2023; Schwarzkopf et al., 2011), thus rejecting our preregistered hypothesis I_2.

Delboeuf illusion magnitude also correlated strongly with objective thresholds for discriminating circle stimuli ($r=0.648$, $BF_{+0}=7.136$). We also observed positive correlations between circle discrimination and the Ebbinghaus and Ponzo, the other illusions using circle targets, but statistical evidence for these correlations was inconclusive ($BF_{+0}<3$). Our hypotheses about circle discrimination abilities were therefore only partially supported. Interestingly, we found moderate evidence for the null hypothesis of no relationship between thresholds for line discrimination and all illusion magnitudes ($BF_{+0}<⅓$), except the Delboeuf ($r=-0.006$, $BF_{+0}=0.349$). This means not even the Mueller-Lyer illusion correlated with this objective task, despite using the same line targets, which rejects our preregistered hypothesis.

There was also evidence against any relationship between the Mueller-Lyer and the Ebbinghaus illusion ($r=-0.334$, $BF_{+0}=0.192$). As such, our results do not adjudicate between the two exploratory hypotheses we specified for the Mueller-Lyer in that it was neither linked with the assumed low-level effects (Ebbinghaus, Delboeuf) nor the presumably higher-level Ponzo illusion. Finally, discrimination thresholds for lines and circles were effectively uncorrelated, although statistical evidence for the alternative or null hypothesis was inconclusive ($r=0.101$, $BF_{+0}=0.455$).

Task modulates illusion magnitude

We next ran a battery of Ebbinghaus illusion measurements, using an Adjustment task, an Unconstrained-2AFC task (unlimited stimulus duration and free eye gaze), and a Fast-2AFC task (brief stimulus duration and instruction to maintain fixation). In this experiment, we also compared two different stimulus designs for the illusion (Figure 2A-B). Averaged across observers, illusion magnitudes varied substantially between the three tasks (Figure 2C; effect of task: BF_{incl}=7.143), confirming our preregistered hypothesis H_1. Specifically, illusion magnitudes were lower when measured with the Adjustment task than either the Unconstrained-2AFC (BF_{10}=155.604) or the Fast-2AFC task (BF_{10}=16.919). However, average illusion magnitude was similar for the two 2AFC tasks although evidence for this null effect was inconclusive (BF_{10}=0.414), leaving hypothesis H_{1b} undetermined.

To further explore these data, we also compared discrimination thresholds for the 2AFC tasks (Figure 2D). Thresholds were consistently smaller for the Unconstrained-2AFC than the Fast-2AFC task (BF_{10}=67.82), meaning the latter was more difficult.

Different stimulus parameters do not affect illusion magnitude

While there was a hint that the stimuli with White&Small circles yielded the weaker illusions, at least for the two tasks with unlimited stimulus duration and free eye gaze, statistical support for any differences between stimulus parameters was inconclusive (BF_{incl}=0.398), rejecting our preregistered hypothesis. Importantly, this applies to the relative illusion effect, expressed as the ratio between test and reference stimulus diameters on a logarithmic scale. In terms of absolute size differences in degrees of visual angle, the illusion magnitude for the White&Small stimuli would be approximately half that of the Grey&Large stimuli.

The lack of any difference in illusion magnitude stands in contrast to a pronounced differences in discrimination ability. Thresholds were consistently worse (larger) for White&Small than Grey&Large circles (Figure 2D; BF_{10}=18.448). Especially for the Fast-2AFC task, judgements became difficult for White&Small circles, to the extent that we decided to deviate from our preregistered exclusion criteria: the discrimination bandwidth for one observer exceeded our predefined criterion for this condition but we decided to include them as this pattern was consistent with the whole sample and their other data were of high quality.

Illusion magnitudes for the two stimulus conditions were also strongly correlated, both averaged across tasks (r=0.873, BF_{+0}=158.861) as well as separately for the Adjustment (r=0.918, BF_{+0}=697.718) and the Unconstrained-2AFC task (r=0.926, BF_{+0}=996.191). For the Fast-2AFC task illusion magnitudes were also positively correlated but statistical evidence for this relationship was not conclusive (r=0.528, BF_{+0}=2.416). Exploring this further we found that the correlation between stimulus conditions was significantly lower for the Fast-2AFC task than both the Adjustment (δ_R=0.389, 5th percentile: 0.037) and the Unconstrained-2AFC task (δ_R=0.397, 5th percentile: 0.128).

Brief stimulation and fixation affect illusion measurements

We then calculated the correlations between the illusion magnitudes from the different tasks and stimulus conditions. This revealed that the Adjustment and Unconstrained-2AFC tasks were strongly correlated with one another (Figure 2E; all BF_{+0}>18.394). In contrast, while the Fast-2AFC task also correlated positively with the Adjustment task, these correlations were significantly attenuated (Grey&Large: δ_T=-2.591, 95th percentile: -1.287; White&Small: δ_T=-1.966, 95th percentile: -0.488). This fully supports our preregistered hypotheses about the individual differences between tasks. We observed the weakest correlation between the Fast-2AFC task with White&Small stimuli and the other conditions. While positive and evidence tended to be in favour of supporting the presence of correlations (all BF_{+0}>1), only some of the correlations for Grey&Large stimuli (with Adjustment: r=0.662, BF_{+0}=6.443; with Unconstrained-2AFC: r=0.734, BF_{+0}=13.581) and one for White&Small stimuli (with Unconstrained-2AFC with grey and small stimuli: r=0.626, BF_{+0}=4.75) conclusively supported this relationship with the other tasks.

Delboeuf illusion shares most of its variance with Ebbinghaus

These lower correlations for White&Small stimuli could be due to overall poorer performance for this task. However, the illusion measurements for the Grey&Large stimuli were highly reliable. This was effectively the same condition as the Ebbinghaus task the first experiment, using the same observers (except for one who was excluded based on preregistered criteria in the second experiment). Comparing these measurements revealed a strong correlation between illusion magnitudes (Figure 2F, black curve; r=0.788, BF_{+0}=28.569) and a slightly weaker correlation between discrimination bandwidths (Figure 2F, blue curve; r=0.704, BF_{+0}=9.718). This test-retest reliability for illusion magnitude is only slightly greater than the correlation between the Ebbinghaus and Delboeuf illusions (Figure 1H; r=0.716).

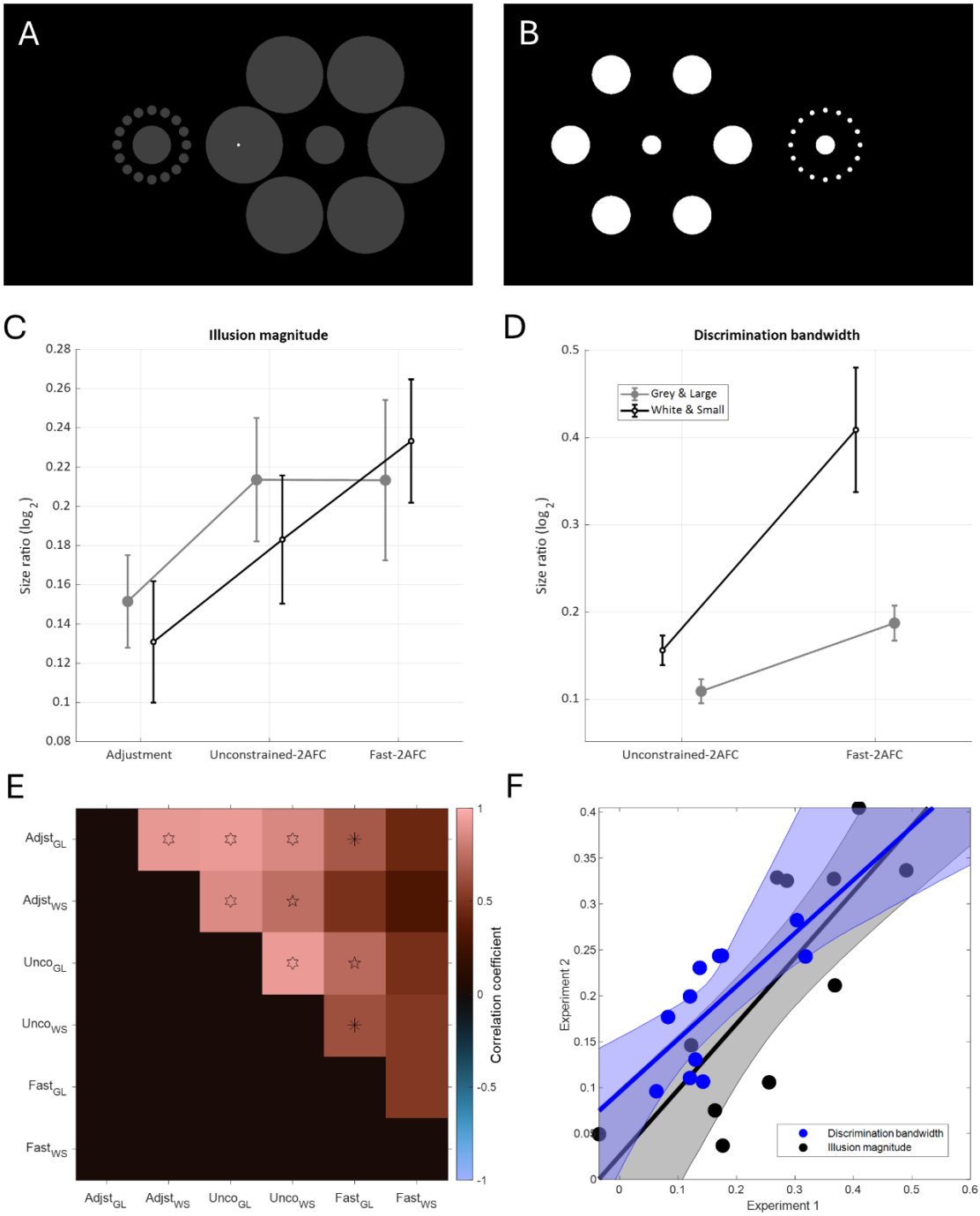

Figure 2. Stimuli (A-B) and results (D-F) of Experiment 2. We tested two configurations of the Ebbinghaus illusion: our standard stimulus from Experiment 1 (A) or another version with all circles half the diameter and in white (B). In separate experimental runs, we used an Adjustment, Unconstrained-2AFC, or a Fast-2AFC task to estimated the illusion magnitude and discrimination sensitivity (except for Adjustment task). Average estimates for illusion magnitude (C) and discrimination bandwidths (D) for each task. Large grey discs denote the Grey&Large stimulus condition and small unfilled discs the White&Small condition. Error bars denote ±1 standard error of the mean across observers. E. Correlation matrix comparing the magnitudes estimated for each condition (task and stimulus). Adjst: Adjustment task, Unco: Unconstrained-2AFC, Fast: Fast-2AFC. GL: Grey&Large, WS: White&Small. Colours indicate strength of correlation. In each cell, symbols denote Bayes Factors: asterisk: $BF_{+0}>3$, pentagram: $BF_{+0}>10$, hexagram: $BF_{+0}>100$, circles $BF_{+0}<⅓$. F. Illusion magnitude (black) and discrimination bandwidth (blue) estimated for the Fast-2AFC Grey&Large condition in Experiment 2 against the corresponding estimates in Experiment 1. Each dot denotes an individual observer. Solid lines show the linear regression fit and the shaded regions its bootstrapped 95% confidence interval.

Discussion

Our findings demonstrate a strong correlation between the magnitudes of the Ebbinghaus and Delboeuf illusions when using a classical two-alternative forced choice procedure, as we used in our earlier work investigating the links between perception and neural processing. This correlation is almost as strong as the test-retest correlation for the Ebbinghaus task itself. These findings confirm previous suggestions that the Ebbinghaus and Delboeuf are essentially manifestations of the same underlying effect (Roberts et al., 2005). Even though Ebbinghaus stimuli are considerably more complex, with multiple contextual stimuli defined by numerous parameters that likely interact with one another as well as with the target, the Delboeuf illusion magnitude remarkably explains most of the variance in the Ebbinghaus.

This finding however contrasts with other reports implying that these two illusions are not substantially related (Jastrzębowska et al., 2023). From our present findings, we deduce that this discrepancy could be due to the methodology used in those studies: previous research mostly used adjustment procedures that not only fail to control stimulus duration and retinotopic location but are also likely skewed by cognitive and/or decision-making biases (M. J. Morgan et al., 2013). Even the few cases where previous studies employed 2AFC procedures, these experiments did not seem to control stimulus characteristics and were also likely susceptible to decision confounds. In fact, the correlation between tasks in that earlier study was considerably weaker than what we found comparing Adjustment and Unconstrained-2AFC tasks in our second experiment. This could call the reliability of these earlier experiments into question. One explanation is that we used more trials for our Adjustment procedure and therefore could estimate the illusion magnitude more precisely.

Our present study took care to match the stimulus conditions between the illusions and discrimination tasks. This is essential for drawing conclusions about how these perceptual effects relate to neural representations in visual cortex and how they may depend on local cortical magnification. Importantly, this also maximises the chance of observing any correlation between the illusions; all else being equal, the only remaining difference is the illusion used. Under the assumption that these effects depend on visual cortex representations, when the conditions differ in terms of geometric dimensions and where in the visual field the stimuli are presented, any relationship between them must be diminished.

As such, it is perhaps not that unsurprising we also found some less pronounced correlations between the Ponzo illusion and the other two using circle targets (Ebbinghaus, Delboeuf) and that all these illusions correlated positively (albeit only relatively weakly and mostly without being statistically conclusive) with the objective ability to discriminate circle widths. While there have been reports of an association between the Ebbinghaus and Ponzo (Grzeczkowski et al., 2017), other research found at best weak association between these illusions (Jastrzębowska et al., 2023; Schwarzkopf et al., 2011), consistent with the idea that the Ebbinghaus is largely mediated by local interactions between V1 neurons while the Ponzo involves interpretations of the three-dimensional scene, a process that must recruit higher-

order processes (Chen et al., 2018; Song et al., 2011). Our findings now add further nuance to this interpretation. It is possible that the Ponzo represents a mixture of both low- and high-level processes. The correlation captures the former variance, which is shared with the Ebbinghaus and Delboeuf, but there also remains unexplained variance stemming from higher-order mechanisms.

Given our matching stimulus characteristics, the most striking aspect about our results is in fact the *absence* of positive correlations, supported by statistical evidence, most notably when comparing the Mueller-Lyer illusion with the others. The neural mechanisms mediating the Mueller-Lyer remain unclear, as it could involve both high-level interpretation of three-dimensional perspective (Gregory, 1966) and low-level image filtering (Carrasco et al., 1986; Ho & Schwarzkopf, 2022). Our results suggest that if the Ponzo and the Mueller-Lyer both involve higher-level processes, these are unlikely to be the same mechanism. Curiously, the Mueller-Lyer also did not correlate with the objective ability to discriminate line lengths, and the two objective discrimination tasks (line length and circle width) also appear unrelated. This indicates that the perceptual judgement in the Mueller-Lyer is rather different from that mediating the other tasks. The Mueller-Lyer is also by far the strongest and most consistent of the illusions we tested. In future work, it could be interesting to design a Mueller-Lyer illusion width circle targets (e.g., surrounded by a crown of arrowheads) to ascertain if the absence of correlation we observed here is due to the different target stimuli.

In our second experiment, we compared Ebbinghaus illusion measurements between different tasks. We found that estimates from Adjustment tasks were very similar to those obtained with the Unconstrained-2AFC task, where observers could inspect stimuli for as long as they needed and were free to move their eyes. In contrast, estimates were far less consistent between these tasks and the Fast-2AFC task, where eye gaze and stimulus duration were tightly controlled. It is tempting to argue that these diminished correlations are a trivial consequence of poorer discrimination performance for the Fast-2AFC task. Indeed, discrimination thresholds for this task, and especially the stimuli comprising White&Small circles, were consistently worse. However, Ebbinghaus illusion measurements were highly reliable and revealed strong correlations with the Delboeuf and Ponzo illusions in Experiment 1. This again stands in contrast to previous studies using Adjustment tasks which failed to find any compelling associations between these illusions (Jastrzębowska et al., 2023). Therefore, we surmise that the Fast-2AFC task can yield robust measurements. The lower correlation between the Fast-2AFC task and Adjustment estimates, compared to the Unconstrained-2AFC task, likely results from the latter's shared critical characteristics with Adjustment: unlimited stimulus presentation and unconstrained eye movements.

This means that the Fast-2AFC design must drive different neuronal populations in visual cortex than the Adjustment and Unconstrained-2AFC tasks. Only the Fast-2AFC task allows relatively tight control over this aspect. It is difficult to ascertain which parts of the visual field – and therefore which corresponding neuronal populations – observers use for their perceptual judgements in the Adjustment and Unconstrained-2AFC tasks. Whenever an

observer foveates one of the two target stimuli, this stimulus will be represented by the highest spatial resolution in foveal cortex. Meanwhile, the other stimulus is in a peripheral location represented with considerably reduced resolution. The eccentricity of this fellow stimulus would then be double that for both, the test and reference stimuli, in the Fast-2AFC task. If the cortical representation matters to the perceptual effect, the illusion estimate must inevitably differ. Notably, stimuli are typically perceived as smaller in the periphery (Anstis, 1998; Moutsiana et al., 2016; Newsome, 1972; von Helmholtz, 1924). Because observers fixate on the midpoint between the two stimuli in the Fast-2AFC, this task controls for this aspect; in contrast, the illusion estimates in the other two tasks will include this general perceptual bias. The situation is complicated by the fact that observers can alternately foveate the two targets, and they may also occasionally fixate in between them. It is difficult to determine which of these states contribute most to the illusion estimate. Either way, it also seems likely that these unconstrained, unlimited tasks leave more room for decision factors to influence the response.

In that light, it is puzzling that previous studies using adjustment tasks found so few correlations between the different illusions (Axelrod et al., 2017; Jastrzębowska et al., 2023). It would be a reasonable assumption that if these tasks were dominated by decision factors, they should be more consistent with each other. For example, if this reflected the conservativism of their decision criterion to reproduce a perceptual match, why should this differ so wildly between these illusions? Such a decision criterion could be considered a common factor to these illusions – the data across the literature speak for themselves that such a factor does not exist, or if it does it only explains a very small amount of variance and could not be reliably detected even in large-scale studies.

Certainly, such a decision factor should not be reflected by the functional architecture in V1. Rather, local properties of V1, which vary with visual field location and geometric stimulus dimensions, will exert specific effects on perception. This does not mean illusion estimates do not generalise up to a point across stimulus characteristics and visual field locations – they should because neural properties also generalise. The cortical magnification factor or population receptive field sizes at 3° eccentricity must at least correlate somewhat with those at 1° eccentricity. Indeed, as we found in the present study, the illusion estimates for the two stimulus conditions in the second experiment were very similar. More generally, even though there was a significant difference between the tasks, the illusion estimates were still relatively well correlated across tasks. Up to a point it should be possible to use comparable stimuli and experimental parameters in studies of perception and how it relates to the brain. However, the more differences are introduced into the design, the greater the variability becomes and the less likely it is that two measures correlate.

All this again highlights a critical fact already discussed: any illusion estimate must reflect multiple separate mechanisms (some of which may interact), ranging from basic perceptual processing to complex decision-making. Therefore, any notion of a single common factor that could be shared by all illusions, even only all spatial contextual illusions of size as studied here,

is fundamentally misguided. Even for potential common factors like general intelligence or personality traits, extraneous variables can influence the outcome of specific tasks in individuals – for example, a colour-blind person would find it difficult to pick a correct target defined by colours even if intelligence is not the limiting factor. However, this does not apply to visual illusions. A common factor, equivalent to general intelligence, that determines a person's susceptibility to visual illusions is based on circular logic. To know that your percept disagrees with the physical reality of the stimulus requires the ability to separate the perceived stimulus from the retinal image. It seems inconceivable that such a process would have evolved, especially because these illusions presumably reflect stimulus processing necessary for functional vision.

However, the opposite implication suggested by several studies (Cretenoud et al., 2021; Grzeczkowski et al., 2017; Jastrzębowska et al., 2023) is even more implausible, namely that there are no factors shared at least by some illusions at all. Taken to its logical conclusion, this notion implies that each illusion is mediated by entirely separate neural mechanisms. This is akin to the concept of a "grandmother neuron" in memory. Just as it seems unlikely that there is only one cell (or a small population of cells) that encodes the identify of a specific person, it is nonsensical to assume that there are specific neural mechanisms for the Ebbinghaus, Delboeuf, Ponzo, or Mueller-Lyer illusions. Many visual illusions are discovered by accident or deliberately designed based on prior knowledge of how visual processing works. They tap into mechanisms. Therefore, illusions which partly tap into the same mechanisms must also be statistically related to each other. More research is required to identify these factors.

Data availability

All stimulus presentation materials, data, and analysis code are publicly available: https://osf.io/5zs7y and https://osf.io/b3h4x

Acknowledgements

We thank Adina Giurgiu and Arash Chaparian for their help in setting up our new testing room.

References

Altan, E., Boyaci, H., & Schwarzkopf, D. S. (2023). Evidence for high-level processing in a Ponzo-like size illusion. *Journal of Vision*, *23*, 4715. https://doi.org/10.1167/jov.23.9.4715

Anstis, S. (1998). Picturing peripheral acuity. *Perception*, *27*(7), 817–825.

Axelrod, V., Schwarzkopf, D. S., Gilaie-Dotan, S., & Rees, G. (2017). Perceptual similarity and the neural correlates of geometrical illusions in human brain structure. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/srep39968

Bosten, J. M., & Mollon, J. D. (2010). Is there a general trait of susceptibility to simultaneous contrast? *Vision Research*, *50*(17), 1656–1664. https://doi.org/10.1016/j.visres.2010.05.012

Brislin, R. W. (1974). The Ponzo Illusion: Additional Cues, Age, Orientation, and Culture. *Journal of Cross-Cultural Psychology*, *5*(2), 139–161. https://doi.org/10.1177/002202217400500201

Carrasco, M., Figueroa, J. G., & Willen, J. D. (1986). A test of the spatial-frequency explanation of the Müller-Lyer illusion. *Perception*, *15*(5), 553–562. https://doi.org/10.1068/p150553

Chen, L., Qiao, C., Wang, Y., & Jiang, Y. (2018). Subconscious processing reveals dissociable contextual modulations of visual size perception. *Cognition*, *180*, 259–267. https://doi.org/10.1016/j.cognition.2018.07.014

Choplin, J. M., & Medin, D. L. (1999). Similarity of the perimeters in the Ebbinghaus illusion. *Perception & Psychophysics*, *61*(1), 3–12.

Coren, S. (1970). Lateral inhibition and geometric illusions. *The Quarterly Journal of Experimental Psychology*, *22*(2), 274–278. https://doi.org/10.1080/00335557043000212

Coren, S., & Enns, J. T. (1993). Size contrast as a function of conceptual similarity between test and inducers. *Perception & Psychophysics*, *54*(5), 579–588.

Cretenoud, A. F., Francis, G., & Herzog, M. H. (2020). When illusions merge. *Journal of Vision*, *20*(8), 12–12. https://doi.org/10.1167/jov.20.8.12

Cretenoud, A. F., Grzeczkowski, 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), 4. https://doi.org/10.1167/jov.20.6.4

Cretenoud, A. F., Grzeczkowski, 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. https://doi.org/10.1167/jov.21.5.26

Deni, J. R., & Brigner, W. L. (1997). Ebbinghaus illusion: Effect of figural similarity upon magnitude of illusion when context elements are equal in perceived size. *Perceptual and Motor Skills*, *84*(3 Pt 2), 1171–1175.

Fang, F., Boyaci, H., Kersten, D., & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. *Current Biology*, *18*(21), 1707–1712. https://doi.org/10.1016/j.cub.2008.09.025

Finlayson, N. J., Papageorgiou, A., & Schwarzkopf, D. S. (2017). A new method for mapping perceptual biases across visual space. *Journal of Vision*, *17*(9), 5. https://doi.org/10.1167/17.9.5

Gallagher, R. M., Suddendorf, T., & Arnold, D. H. (2019). Confidence as a diagnostic tool for perceptual aftereffects. *Scientific Reports*, *9*(1), 7124. https://doi.org/10.1038/s41598-019-43170-1

Gregory, R. L. (1966). *Eye and brain: The psychology of seeing*. McGraw-Hill.

Grzeczkowski, 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

He, D., Mo, C., Wang, Y., & Fang, F. (2015). Position shifts of fMRI-based population receptive fields in human visual cortex induced by Ponzo illusion. *Experimental Brain Research*. https://doi.org/10.1007/s00221-015-4425-3

Ho, M.-L., & Schwarzkopf, D. S. (2022). The human primary visual cortex (V1) encodes the perceived position of static but not moving objects. *Communications Biology*, *5*, Article 1. https://doi.org/10.1038/s42003-022-03136-y

Howell, D. C. (1997). *Statistical Methods for Psychology*. Duxbury Press.

JASP Team. (2023). *JASP (Version 0.18.3)* [Computer software]. https://jasp-stats.org

Jastrzębowska, M. A., Ozkirli, A., Cretenoud, A. F., Draganski, B., & Herzog, M. H. (2023). *Is there a neural common factor for visual illusions?* (p. 2023.12.27.573437). bioRxiv. https://doi.org/10.1101/2023.12.27.573437

Jogan, M., & Stocker, A. A. (2014). A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. *Journal of Vision*, *14*(3), 20. https://doi.org/10.1167/14.3.20

Mareschal, I., Morgan, M. J., & Solomon, J. A. (2010). Cortical distance determines whether flankers cause crowding or the tilt illusion. *Journal of Vision*, *10*(8), 13. https://doi.org/10.1167/10.8.13

Morgan, M., Dillenburger, B., Raphael, S., & Solomon, J. A. (2012). Observers can voluntarily shift their psychometric functions without losing sensitivity. *Attention, Perception & Psychophysics*, *74*(1), 185–193. https://doi.org/10.3758/s13414-011-0222-7

Morgan, M. J., & Glennerster, A. (1991). Efficiency of locating centres of dot-clusters by human observers. *Vision Research*, *31*(12), 2075–2083. https://doi.org/10.1016/0042-6989(91)90165-2

Morgan, M. J., Melmoth, D., & Solomon, J. A. (2013). Linking hypotheses underlying Class A and Class B methods. *Visual Neuroscience*, *30*(5–6), 197–206. https://doi.org/10.1017/S095252381300045X

Moutsiana, C., de Haas, B., Papageorgiou, A., van Dijk, J. A., Balraj, A., Greenwood, J. A., & Schwarzkopf, D. S. (2016). Cortical idiosyncrasies predict the perception of object size. *Nature Communications*, *7*, 12110. https://doi.org/10.1038/ncomms12110

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, *9*(3), 429–434. https://doi.org/10.1038/nn1641

Newsome, L. R. (1972). Visual angle and apparent size of objects in peripheral vision. *Perception & Psychophysics*, *12*(3), 300–304. https://doi.org/10.3758/BF03207209

Ni, A. M., Murray, S. O., & Horwitz, G. D. (2014). Object-Centered Shifts of Receptive Field Positions in Monkey Primary Visual Cortex. *Current Biology: CB*. https://doi.org/10.1016/j.cub.2014.06.003

Patten, M. L., & Clifford, C. W. G. (2015). A bias-free measure of the tilt illusion. *Journal of Vision*, *15*(15), 8. https://doi.org/10.1167/15.15.8

Poom, L. (2020). Influences of orientation on the Ponzo, contrast, and Craik-O'Brien-Cornsweet illusions. *Attention, Perception, & Psychophysics*, *82*(4), 1896–1911. https://doi.org/10.3758/s13414-019-01953-8

Pooresmaeili, A., Arrighi, R., Biagi, L., & Morrone, M. C. (2013). Blood Oxygen Level-Dependent Activation of the Primary Visual Cortex Predicts Size Adaptation Illusion. *Journal of Neuroscience*, *33*(40), 15999–16008. https://doi.org/10.1523/JNEUROSCI.1770-13.2013

Roberts, B., Harris, M. G., & Yates, T. A. (2005). The roles of inducer size and distance in the Ebbinghaus illusion (Titchener circles). *Perception*, *34*(7), 847–856.

Sánchez-Fuenzalida, N., van Gaal, S., Fleming, S. M., Haaf, J. M., & Fahrenfort, J. J. (2023). Predictions and rewards affect decision-making but not subjective experience.

Proceedings of the National Academy of Sciences of the United States of America, *120*(44), e2220749120. https://doi.org/10.1073/pnas.2220749120

Schwarzkopf, D. S., & Rees, G. (2013). Subjective size perception depends on central visual cortical magnification in human v1. *PloS One*, *8*(3), e60550. https://doi.org/10.1371/journal.pone.0060550

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

Song, C., Schwarzkopf, D. S., Kanai, R., & Rees, G. (2015). Neural Population Tuning Links Visual Cortical Anatomy to Human Visual Perception. *Neuron*, *85*(3), 641–656. https://doi.org/10.1016/j.neuron.2014.12.041

Song, C., Schwarzkopf, D. S., Lutti, A., Li, B., Kanai, R., & Rees, G. (2013). Effective connectivity within human primary visual cortex predicts interindividual diversity in illusory perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(48), 18781–18791. https://doi.org/10.1523/JNEUROSCI.4201-12.2013

Song, C., Schwarzkopf, D. S., & Rees, G. (2011). Interocular induction of illusory size perception. *BMC Neuroscience*, *12*, 27. https://doi.org/10.1186/1471-2202-12-27

Song, C., Schwarzkopf, D. S., & Rees, G. (2013). Variability in visual cortex size reflects tradeoff between local orientation sensitivity and global orientation modulation. *Nature Communications*, *4*, 2201. https://doi.org/10.1038/ncomms3201

Sperandio, I., Chouinard, P. A., & Goodale, M. A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nature Neuroscience*, *15*(4), 540–542. https://doi.org/10.1038/nn.3069

Urale, P. W. B., & Schwarzkopf, D. S. (2023). Effects of cortical distance on the Ebbinghaus and Delboeuf illusions. *Perception*, *52*(7), 459–483. https://doi.org/10.1177/03010066231175014

von Helmholtz, H. (1924). *Helmholtz's treatise on physiological optics, Vol. 1, Trans. From the 3rd German ed* (pp. xxi, 482). Optical Society of America. https://doi.org/10.1037/13536-000

Yildiz, G. Y., Sperandio, I., Kettle, C., & Chouinard, P. A. (2021). The conceptual understanding of depth rather than the low-level processing of spatial frequencies drives the corridor illusion. *Vision Research*, *181*, 21–31. https://doi.org/10.1016/j.visres.2020.11.012