Evaluating Graphical Perception in Virtual Reality: Evidence from Two Visualization Experiments

Lucas Siqueira Rodrigues
lucas.rodrigues@tum.de
Technical University of Munich
Munich, Germany

Johann Habakuk Israel
israel@htw-berlin.de
HTW Berlin
Berlin, Germany

Thomas Kosch
thomas.kosch@hu-berlin.de
HU Berlin
Berlin, Germany

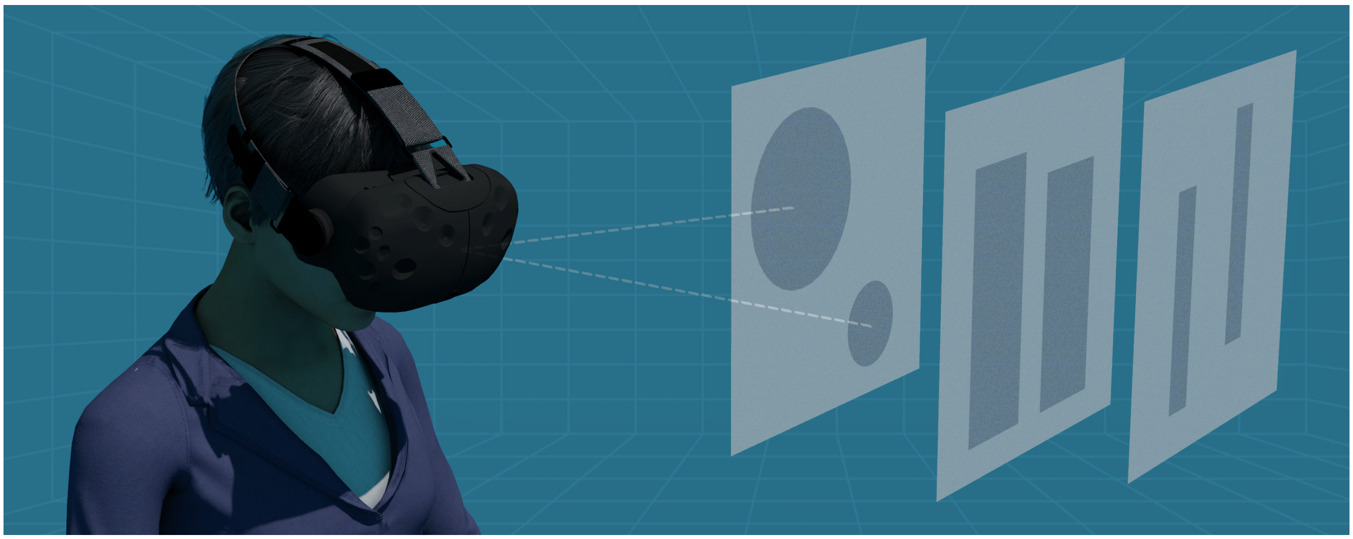

Figure 1: We examine two-dimensional graphical perception in VR across virtual tabletop and wall display conditions, mirroring foundational graphical perception studies. In Experiment 1, bar charts supported the most accurate judgments, followed by length and area encodings. In Experiment 2, accuracy degraded significantly with viewing distance for horizontal bar charts.

Abstract

Although Virtual Reality (VR) is increasingly used for information work, there is a limited understanding of how data should be encoded in immersive visualizations to support accurate interpretation. It remains unclear whether established two-dimensional graphical perception guidelines generalize to VR settings with different viewing conditions. To address this gap, we conducted two immersive experiments in which participants performed ratio-estimation tasks, with mean absolute error as the metric. In Experiment 1, we replicated a foundational graphical perception study on a virtual tabletop display and found that accuracy followed the established ranking of visual variables, with the lowest mean absolute error for bar charts, followed by length and area encodings. In Experiment 2, we examined the effects of chart orientation and viewing distance on a virtual wall display, extending an established large-display paradigm to VR, and found that horizontal bar charts produced higher error and greater between-participant variability than vertical bar charts, with accuracy degrading more sharply with viewing distance for horizontal than vertical orientations. These findings indicate that while classic perceptual rankings largely carry over to VR, viewing conditions interact with chart design in ways that affect accuracy, informing recommendations for optimal encoding of 2D visualizations for immersive environments.

CCS Concepts

• **Human-centered computing → Empirical studies in visualization**; **Visualization theory, concepts and paradigms**; • **Computing methodologies → Perception**.

Keywords

Graphical Perception, Virtual Reality, Information Visualization, Human-Computer Interaction, Psychophysics, Immersive Analytics

ACM Reference Format:
Lucas Siqueira Rodrigues, Johann Habakuk Israel, and Thomas Kosch. 2026. Evaluating Graphical Perception in Virtual Reality: Evidence from Two Visualization Experiments. In *Proceedings of 52nd Graphics Interface Conference (GI '26)*. ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

GI '26, Waterloo, Ontario

ACM ISBN 978-1-4503-XXXX-X/2026/06
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Virtual Reality (VR) has become a platform for consequential information work, enabled by advances in display technology, tracking, and overall accessibility [10, 21]. Major technology platforms are actively positioning immersive environments as general-purpose workspaces, where users attend remote meetings, review documents, and collaborate on shared content [27]. Similar developments have taken place in higher education, where VR is increasingly used to deliver lectures and facilitate collaborative learning across distances [34]. As these use cases grow, the effective presentation of data in immersive environments becomes a practical concern for interface designers.

A natural question is whether the established findings of graphical perception research carry over to VR. Decades of work on 2D displays have produced solid guidelines for which visual encodings support accurate judgments [7, 24, 29, 31]. These findings guide the design of data visualization tools on conventional displays. Whether they generalize to VR, where depth compression, viewing distance, and display orientation differ from desktop conditions, is not understood [9, 26]. Some recent evidence suggests that relative graphical judgments in VR can approximate flat-screen accuracy [12, 37], but systematic evaluations of how graph type, chart orientation, and viewing distance jointly influence perceptual accuracy in immersive settings remain limited. In particular, it is unclear whether the ranking of visual variables established on conventional displays holds in VR, and how orientation and viewing position on wall-sized virtual displays affect estimation performance.

We present two conceptual replications of graphical perception studies conducted in VR. Experiment 1 replicates Cleveland & McGill's comparison of visual variables [7] using a virtual tabletop display, and Experiment 2 replicates Bezerianos & Isenberg's investigation of perception on a large wall display [5] in an immersive setting. We chose to replicate the paradigms of these specific studies to address two structural baselines in data visualization. First, establishing whether Cleveland & McGill's foundational findings hold true under a Head-Mounted Display (HMD) is a prerequisite for adapting 2D visualization paradigms to VR. Second, Bezerianos & Isenberg's large-display framework serves as the most direct physical analogue to the infinite-canvas 'virtual walls' frequently deployed in spatial computing workspaces, allowing us to isolate the effects of viewing distance and orientation. Thirteen participants completed both experiments using a head-mounted display, judging ratios between pairs of graphical stimuli. We assess how perceptual rankings and distance effects manifest in immersive environments and offer recommendations for encoding 2D visualizations in VR contexts that require accurate data interpretation.

This paper makes three contributions:

- Evaluation of generizalibility of established perceptual rankings through a conceptual replication of Cleveland & McGill graphical perception study to VR.
- A systematic analysis of the interaction between chart orientation and viewing distance on perceptual accuracy in large virtual wall displays through a VR extension of Bezerianos & Isenberg.
- Design implications for encoding quantitative information in immersive analytics environments.

2 Related Work

2.1 Graphical Perception

Graphical perception studies how viewers decode data from visual encodings [7]. Cleveland & McGill's classic work showed that position and length judgments produce much lower error than judgments of area or volume [7]. Their ranking of visual channels has guided chart design ever since. Subsequent psychophysical studies on 2D displays have largely confirmed these findings. For example, Talbot *et al.* replicated Cleveland & McGill's bar-chart experiments online and "generally confirmed" the original rankings, providing further insight into why aligned bars beat stacked bars [35]. Simkin and Hastie likewise found that comparing heights along a common baseline (simple bars) is the most accurate judgment [29]. Heer and Bostock used Mechanical Turk to replicate classic proportional-judgment tasks, again matching earlier design guidelines [13]. Spence and others have systematically measured such errors for bar charts, pies, etc [31]. In sum, decades of 2D graphical perception research establish that position/length encodings yield the most accurate comparisons, whereas area encodings incur greater under- and overestimation errors.

2.2 Graphical Perception on 2D Displays

This prior work on 2D displays directly motivated our stimuli and hypotheses. In their pen-and-paper studies, Cleveland & McGill showed that comparisons between aligned, adjacent bars (with a common baseline) yield the smallest error [8]. We thus expect vertical bar charts (with a common baseline) to be interpreted more accurately than equivalent length or area encodings. Consistent with this, Talbot *et al.* reported that performance is substantially worse for misaligned bar charts or charts containing distracting intervening bars [35]. Simkin & Hastie similarly found near-baseline comparisons easiest [29]. Spence assessed the accuracy of judgments for graphs of different dimensions and reported small estimation errors across categories, especially for bars, lines, and pie slices as visual variables [31]. Heer & Bostock's work confirms that these classic results generalize to modern interactive plots [13].

Bezerianos & Isenberg studied perception on large tiled wall displays, showing that viewing angle and display distance significantly affect estimation accuracy on wall-sized graphs [5]. Objects viewed at extreme tilt or from far away can appear distorted, biasing length and area judgments. They recommend positioning charts face-on to the viewer and compensating for distortion in wall-display visualizations. The physical construction of such displays introduces an additional confound: Wallace et al. found that tiled displays are assembled from multiple panels separated by bezels, which create visual discontinuities that affect magnitude judgment [36]. Our Experiment 2 is inspired by Bezerianos & Isenberg's work, as we emulate a large wall display in VR, while the absence of physical bezels eliminates this confound and allows us to examine the effects of orientation and distance more cleanly.

2.3 Graphical Perception in Virtual Reality

Virtual reality introduces new perception effects related to immersion in stereoscopic 3D space. A well-known issue is distance compression: Users underestimate distances and sizes in VR [20, 38].

For example, Armbrüster *et al.* found that users perceive far objects as closer than they are, especially outside arm's reach [2]. Renner *et al.* review many such effects, categorizing them into technical (display FOV, stereo), compositional, and human factors [26]. Cutting analyzed depth cues (occlusion, perspective, binocular disparity, etc.) that underlie this distortion [9]. In practice, even with stereo vision, VR reduces apparent scale and depth for distant objects.

Most VR visualization research has focused on 3D data or spatial tasks. While some work reports benefits of stereo cues for spatial judgments [18], other studies warn that true 3D visualizations can slow or hinder users on retrieval tasks compared to 2D layouts [22, 28], suggesting that depth does not universally benefit perceptual accuracy. This motivates examining 2D encodings in VR separately from 3D visualization research.

Recent work has begun to examine 2D graphical perception specifically in immersive settings. Arjun *et al.* found that bar-height encodings yielded higher accuracy and lower cognitive load than area or orientation encodings in a fully immersive dashboard [1]. Similarly, Quach *et al.* demonstrated that body-anchored bar charts support accurate value estimation and comparison in an immersive environment [25]. Perception of large virtual wall displays has also been studied: participants made fewer errors when reading visual variables on curved or tiled VR walls than on flat arrangements, with performance sometimes exceeding that in real-world settings [12]. Watson *et al.* directly tested bar-chart perception in VR, replicating Cleveland & McGill's paradigm inside an HMD and found that VR's distance compression causes bar heights to be underestimated in absolute terms, but relative percentage judgments remained as accurate as on flat screens [37]. Notably, they confirmed that length comparisons were more accurate than angle comparisons and that the distortions in VR primarily affect absolute scale.

Overall, initial work suggests that standard 2D perceptual rankings may carry over to immersive environments, but perceptual differences introduced by VR, whether driven by resolution, depth cues, environmental context, or navigation, warrant investigating the effects of graph type, orientation, and distance on judgment accuracy. Our work fills this gap by examining these factors, extending earlier findings from physical paper and wall-sized displays.

3 Methodology

We utilized a ratio estimation task in a within-subjects design replicating work by Cleveland & McGill [7] and Bezerianos & Isenberg [5] across two experiments: Experiment 1 presented stimuli horizontally on a virtual table, while Experiment 2 presented stimuli vertically as in a virtual wall display (Figure 2). Both experiments used VR to present stimuli, with participants operating a controller to indicate responses. We assessed absolute error to evaluate performance across visual variables and stimulus placements. Informed by prior literature, we propose the following hypotheses:

- **H1:** Accuracy across visual variables will replicate prior findings, with bar charts yielding the lowest absolute error, followed by length and area encodings.
- **H2:** Horizontal bar charts will produce higher absolute errors and variation between participants than their vertical counterparts.
- **H3:** Absolute error and between-participant variability in bar chart judgments will decrease as charts are positioned closer to the viewer.

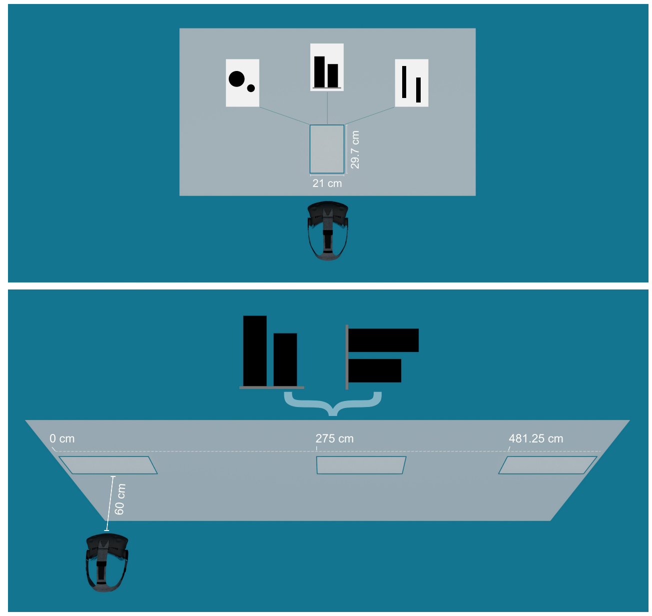

Figure 2: Overview of both experiments. (Top) Experiment 1: Participants judged ratios between pairs of stimuli (circles, lines, bar charts) on a virtual A4 sheet (21 × 29.7 cm). (Bottom) Experiment 2: Participants, seated 60 cm from a virtual wall, judged bar charts with varying orientation (vertical vs horizontal) and horizontal position (0 m, 2.75 m, 4.81 m). Both experiments measured absolute error in ratio estimation.

3.1 Participants

Thirteen participants (students and researchers at Sorbonne Université) completed both experiments. Participants self-reported no visual impairments beyond corrected vision. We recruited participants using convenience sampling through our professional networks, supplemented by snowball sampling via referrals from initial participants. Specific demographic variables are omitted due to an expiration of access to the original survey platform, as elaborated upon in the Limitations section. The order of experiments was counterbalanced across participants to mitigate order effects.

3.2 Procedure

While formal ethics committee approval was not required for this non-clinical study, all procedures involving human participants complied with standard ethical research practices and the 1964 Declaration of Helsinki. Participants were briefed and provided their written informed consent. Then, they adjusted the height of an immobile rotating chair and received instructions to fit and operate the VR HMD and controller. The VR equipment was then calibrated by reconfiguring the tracking center using the SteamVR setup procedure. Participants completed a training session before each experiment and had their questions answered before proceeding. To mitigate systematic learning or fatigue effects, the overall order of the two main experiments (Virtual Table vs. Virtual Wall) was fully counterbalanced across participants. Each participant completed a total of 102 trials, with 54 in the table experiment and 48 in the wall experiment. For Experiment 1, trials were blocked by graph type, and block order was counterbalanced across participants using a balanced Latin square so that each graph type appeared once in each ordinal position and each pair of graph types occurred equally often in each order. For Experiment 2, trials were organized into two orientation-defined blocks (vertical vs. horizontal). Block order was counterbalanced across participants by alternating which orientation was presented first (odd vs. even participant IDs), and within each orientation block, the full set of stimulus combinations (8 ratio levels × 3 positions) was fully randomized for each participant to prevent localized adaptation to a single axis. After the first experiment was conducted, a short break was suggested in accordance with ergonomic guidelines [17, 30].

3.3 Apparatus

We used a desktop computer (Intel Core i5-4570 CPU @ 3.20GHz, 16 GB RAM, NVIDIA Quadro P4000) capable of executing our experimental artifacts at the VR hardware's maximum settings. The application was developed in Unity 3D, targeting Windows and the HTC Vive HMD for stimulus presentation at a 2160 × 1200-pixel resolution, 90 Hz, and 110 degrees Field-of-View (FOV). Participants responded to trials with the standard Vive wand controller, which was used both to indicate ratio estimates via lateral thumbpad movements and to confirm responses via trigger pulls (detailed in Section 3.5.

3.4 Stimuli

Participants viewed pairs of two-dimensional graphical stimuli rendered in VR, appearing as if printed on physical surfaces. In the table experiment, stimulus types included circles, vertical lines, and vertical bar charts, with lines and circles placed at randomized positions to prevent alignment between shapes, while bar charts were aligned at their bases. In this experiment, the larger stimulus randomly varied in size by up to 10% between trials to prevent participants from adapting to a fixed reference size, as trials were blocked by graph type. In the wall experiment, only vertical and horizontal bar charts were used, aligned at their bases, with the larger stimulus kept at a fixed size to replicate the conditions of Bezerianos & Isenberg's original study [5].

3.5 Task

Following Jansen et al. [15], we selected ratio estimation over constant-sum production for its reliability and ease of implementation in VR. On each trial, participants were presented with a pair of stimuli and asked: *"If the larger shape represents 100%, what percentage is the smaller shape?"*. Participants were prompted to provide quick visual judgments rather than precise measurements. Participants used a thumbpad to input integer responses on a 100-point scale (1-100), representing percentage values. The input was polled continuously per frame update, incrementing or decrementing the value by one unit as long as directional input was maintained. Once the desired value was reached, the response was confirmed with a trigger pull. If participants attempted to confirm a response without first adjusting the thumbpad from its default value, a UI alert prompted them to select an answer before proceeding. The response UI was identical across all trials and did not occlude stimuli.

3.6 Experiment 1: Virtual Table

The table experiment was designed to emulate the conditions described by Cleveland & McGill [7]. As illustrated in Figure 3, participants faced a virtual A4 sheet of paper fixed on a standard 72cm tall table. The sheet displayed the trial's stimulus pair alongside a UI for entering and confirming answers, and also served as a canvas for providing participants with initial and final instructions. Similar to Cleveland & McGill, we did not constrain head movements. While Cleveland & McGill randomized the size of the reference stimulus, we fixed it across trials in this experiment to simplify training and reduce participant confusion. Although such invariability could introduce bias, prior work has found a statistically non-significant effect of reference stimulus side [39], supporting this as an informed design deviation. During training, participants were instructed that the larger stimulus was always positioned to the left. We recorded participant answers across the 54 trials of this experiment. Additionally, we tracked task completion time, head position, and head rotation to characterize natural task completion and viewing behavior during tabletop VR interactions.

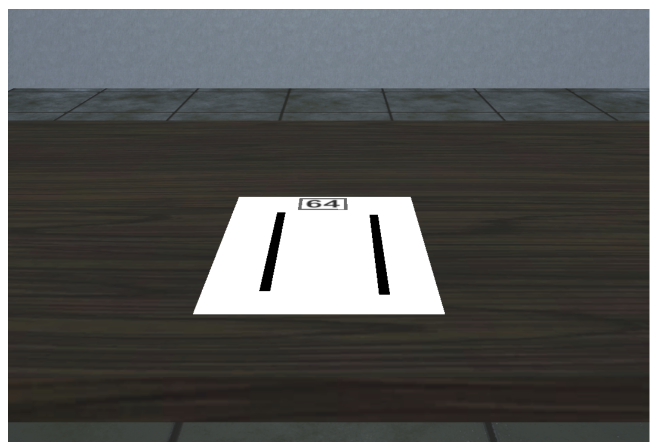

Figure 3: Participant view during a line estimation trial in Experiment 1. Two graphs are displayed on a virtual A4 sheet on the tabletop. The current response value is displayed at the top and adjusted via the controller.

3.7 Experiment 2: Virtual Wall

The wall experiment partially replicated the fixed-position condition of Bezerianos & Isenberg [5], in which participants were positioned 60 cm away from the display. We selected this condition as the original authors reported it yielded the most notable interactions with stimulus placement in their study. Bezerianos & Isenberg reported a significant location × visual variable interaction but no significant effect of screen height. Accordingly, we restricted spatial variations to the horizontal axis and maintained vertical placement aligned to the average participant's eye level. This design choice mirrored the significant dimensions of the original study while preventing the cervical fatigue associated with repetitive vertical head movements (pitch) while wearing a front-heavy HMD across a 102-trial session to maximize statistical power within realistic volunteer time constraints. Figure 4 depicts the average participant's view when facing stimuli placed at their closest position.

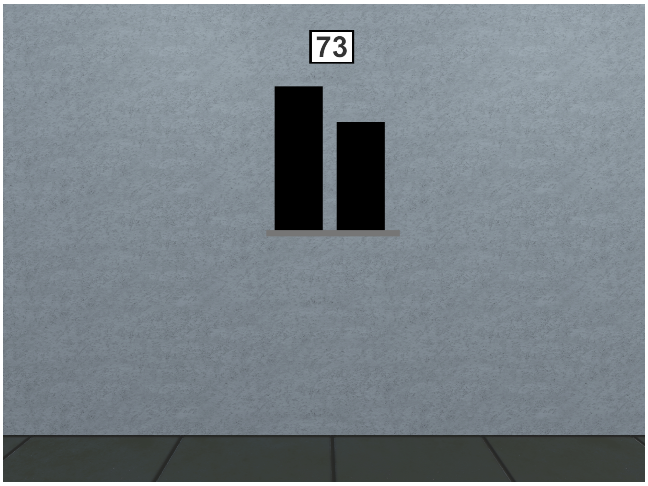

Figure 4: Participant view during a trial in Experiment 2. A vertical bar chart is presented on the textured wall surface at Position 0 cm. The response UI appears above the stimuli.

Bezerianos & Isenberg's original display consisted of multiple LCD monitors with visible bezels, which participants reportedly used as measurement references. Because VR environments lack bezels, this potential confound was not replicated. To provide participants with sufficient depth cues, which Bezerianos & Isenberg's physical wall provided implicitly, we displayed stimuli over a textured wall rather than a black background. Participants were positioned 60 cm from the virtual wall display at its leftmost point, but were free to rotate within the fixed chair to face visualizations. Stimuli pairs were placed at eye level and positioned horizontally at three distances: 0 m, 2.75 m, and 4.8125 m to the participant's right, replicating positions A3, E3, and H3 of Bezerianos & Isenberg's display [5]. The training session for this experiment differed from its table counterpart in that participants were instructed that the larger stimulus was always positioned to the left for vertical bar charts and on top for horizontal bar charts. We recorded participants' answers, the horizontal viewing angle, and task completion time. Because stimuli were sometimes placed outside a participant's initial field of view, we also recorded a secondary completion time from the moment the participant first encountered the stimuli.

3.8 Independent Variables

We define Graph Type as the independent variable for Experiment 1 (Table) and Stimulus Position and Stimulus Orientation as independent variables for Experiment 2 (Wall). Graph Type contains three levels in the table experiment (*Circle*, *Line*, and *Bar Chart*). For the wall experiment, Graph Type was constant as Bar Chart, whereas Stimulus Orientation had two values (*Vertical* and *Horizontal*), while Stimulus Position contained three levels corresponding to horizontal distances of 0 m, 2.75 m, and 4.8125 m from the participant, matching positions A3, E3, and H3 in the aforementioned work of Bezerianos & Isenberg [5].

3.9 Dependent Variable

We measure *Absolute Error* as the dependent variable in both Table and Wall experiments. *Absolute Error* reflects the magnitude of the discrepancy between the ground truth ratio and the participant's ratio estimate. Moreover, we fit psychophysical power functions to characterize systematic perceptual bias in participants' responses.

4 Results

4.1 Results for Experiment 1 - Virtual Table

4.1.1 Absolute Error. We examined absolute error (%) across three visualization conditions (Area, Bar Chart, Length) in a within-subjects experimental design. Figure 5 illustrates the participant-level distributions across conditions. For each participant and condition, we computed the mean absolute error across trials. Shapiro–Wilk tests indicated that aggregated accuracy values did not deviate significantly from normality for any condition (Area: $p = .42$, BarChart: $p = .71$, Length: $p = .083$; all $p > .05$). Because Mauchly's test indicated a violation of sphericity for condition ($W = 0.56$, $p = .043$), we report Greenhouse–Geisser-corrected repeated-measures ANOVA results. In addition, we fitted a linear mixed-effects model with condition as a fixed effect and participant-specific random intercepts to account for all trial-level observations. Pairwise comparisons following the repeated-measures ANOVA were Bonferroni-corrected, while post-hoc comparisons derived from the linear mixed-effects model were Tukey-corrected., and within-subject effect sizes are reported as Cohen's d.

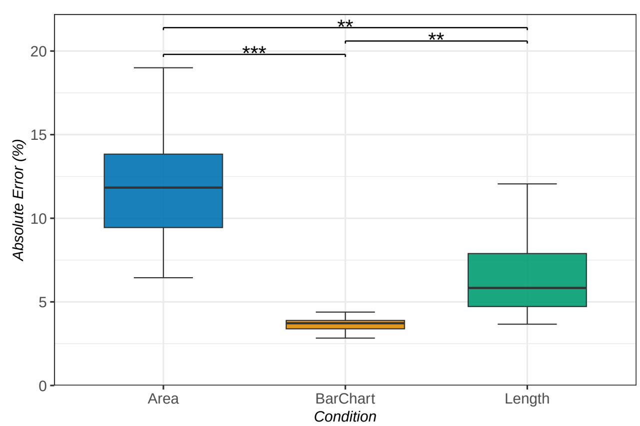

Figure 5: Mean absolute error (|*estimate* − *true ratio*|, percentage points) by visualization condition in Experiment 1. Boxplots summarize participant-level means. Lower values indicate better accuracy. Bar charts yield the lowest error, followed by length and area encodings.

The repeated-measures ANOVA on participant mean accuracy revealed a significant main effect of condition, $F(2, 24) = 33.01$, $p < .001$, Greenhouse–Geisser-corrected $p < .001$, generalized $\eta^2 = .62$. Bonferroni-corrected paired t-tests showed that Area produced significantly higher errors than BarChart ($p < .001$, $d = 2.65$) and Length ($p = .004$, $d = 1.60$), and that Length errors were higher than BarChart errors ($p = .002$, $d = 1.28$). The highest error was observed for Area ($M = 12.05\%$, $SD = 4.03$), followed by Length ($M = 6.61\%$, $SD = 2.60$) and BarChart ($M = 3.68\%$, $SD = 0.69$).

The trial-level linear mixed-effects model with random participant intercepts confirmed this pattern, with BarChart and Length exhibiting substantially lower errors than Area (all Tukey-adjusted $p < .0001$). Area judgments deviated by 12.05 percentage points, compared to 6.61 for Length and 3.68 for BarChart. Bar charts yielded the lowest absolute error, followed by length and area.

For historical context, Figure 6 compares this experiment's estimates with a digitized Cleveland & McGill benchmark for their Type I Bar Chart; this panel is included for descriptive comparison only and was not used for inferential testing.

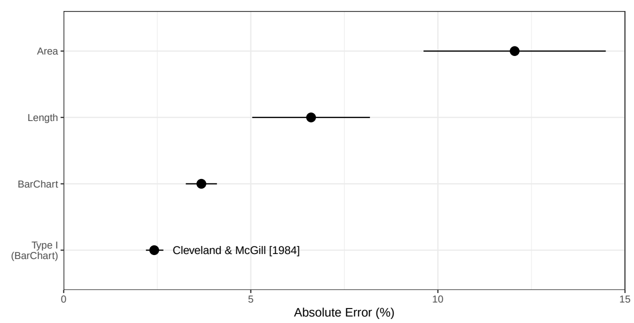

Figure 6: Mean absolute error (|estimate − true ratio|, %) by graph type in Experiment 1. Points show means, and error bars show 95% BCa bootstrap confidence intervals. Bar-charts yielded the lowest error, followed by length and area. Cleveland & McGill (1984) Type I bar chart is added as a reference.

4.1.2 *Psychophysical coefficients (individual response curves).*

As an exploratory and purely descriptive analysis intended to visually characterize systematic bias rather than for formal hypothesis testing, we fitted a psychophysical power function of the form $P = bS^a$ (following Stevens' power law [32]) to each participant's data in each condition (see Figure 7). The exponent a captures how perceived ratios scale with actual ratios, with $a = 1$ indicating veridical (linear) perception, $a > 1$ overestimation, and $a < 1$ underestimation. This analysis is descriptive and intended to assess systematic bias and between-participant variability.

Bar-chart perception showed low between-participant variation (mean $a = 1.04$, $SD = 0.05$). Length estimations showed greater deviation from linearity (mean $a = 1.10$, $SD = 0.15$), indicating modest overestimation and increased variability. Variation was highest for area stimuli (mean $a = 1.49$, $SD = 0.35$), reflecting stronger systematic bias and greater heterogeneity across participants.

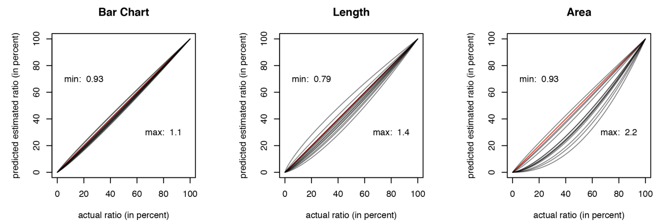

Figure 7: Psychophysical power-function coefficients (accuracy ratios) by stimulus type, summarizing bias and between-participant variability. Values near 1 indicate near-linearity. Bar charts were closest to linearity, length showed modest overestimation and greater spread, and area showed the largest bias and heterogeneity, matching prior findings.

4.2 Results for Experiment 2 - Wall Display

4.2.1 *Absolute Error.*

We analyzed absolute error (%) in a within-subjects 2×3 design with factors Orientation (horizontal, vertical) and Position (0, 1, 2). For each participant, the mean absolute error was computed for each Orientation × Position condition.

Shapiro–Wilk tests on participant-level cell means were non-significant in all six cells ($p \geq .082$), indicating no clear departure from normality. Mauchly's tests indicated no violation of sphericity for Position ($W = 0.993, p = .962$) or the Orientation × Position interaction ($W = 0.852, p = .415$).

A two-way repeated-measures ANOVA showed significant main effects of Orientation, $F(1, 12) = 77.43, p < .001, \eta^2_G = .49$, and Position, $F(2, 24) = 41.38, p < .001, \eta^2_G = .39$, qualified by a significant Orientation × Position interaction, $F(2, 24) = 27.45, p < .001, \eta^2_G = .40$ (Figure 8). Given the significant interaction, simple-effects tests were prioritized for interpretation.

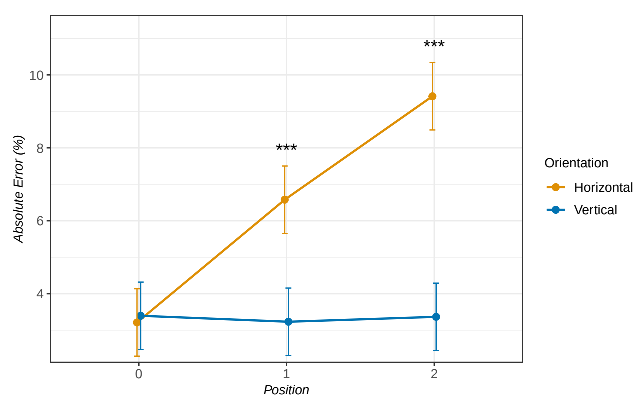

Figure 8: Orientation × Position interaction for absolute error (%) using model-based estimated marginal means from the trial-level mixed-effects model. Orientation differences are negligible at Position 0 but increase at Positions 1 and 2. Asterisks indicate significant horizontal-versus-vertical simple effects within position (*, $p < .001$).**

Orientation differences depended on Position. At Position 0, horizontal and vertical bars did not differ significantly ($p = .713$); at

Positions 1 and 2, horizontal bars yielded larger error than vertical bars (both $p < .001$). In averaged analyses, horizontal bars still showed higher error than vertical bars, paired $t(12) = 8.80, p < .001$, mean difference = 3.07 percentage points, 95% CI [2.31, 3.83], $d = 2.46$. Trial-level mixed-effects estimates showed the same overall pattern (horizontal: 6.40%, vertical: 3.33%; Tukey-adjusted $p < .001$). Thus, H2 is supported, with the strongest orientation differences emerging at mid and far positions.

Position differences were orientation-dependent. Within the horizontal condition, all pairwise position contrasts were significant and large: Position 0 vs 1 (mean difference = 3.37 pp, 95% CI [2.04, 4.69], $p < .001$, $d = 1.91$), Position 0 vs 2 (mean difference = 6.20 pp, 95% CI [4.69, 7.71], $p < .001$, $d = 3.28$), and Position 1 vs 2 (mean difference = 2.84 pp, 95% CI [1.40, 4.27], $p = .001$, $d = 1.22$), indicating monotonically increasing error with distance. Within the vertical condition, no pairwise position contrast was significant ($p = .942, .998, .960$). Collapsed analyses across orientation showed a significant Position effect, $F(2, 24) = 41.38, p < .001, \eta^2_G = .53$, with significant Bonferroni-corrected pairwise differences (0 vs 1: $p = .0012$, $d = 1.20$; 0 vs 2: $p < .001$, $d = 2.83$; 1 vs 2: $p = .0036$, $d = 1.07$). Therefore, H3 is only supported for horizontal bars.

Variability analyses matched this pattern (see Figure 9). Participant-level variability (SD across positions) was greater for horizontal than vertical charts, paired $t(12) = 6.22, p < .001$, mean difference = 2.39, 95% CI [1.55, 3.23], $d = 2.61$. Variability across positions (SD across orientations) also differed by position, $F(2, 24) = 21.85, p < .001, \eta^2_G = .50$, with all Bonferroni pairwise comparisons significant (0 vs 1: $p = .003$; 0 vs 2: $p < .001$; 1 vs 2: $p = .018$).

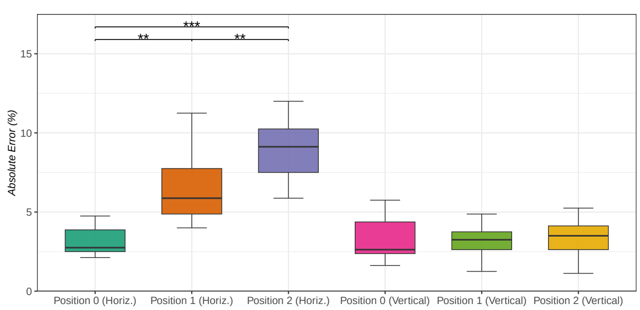

Figure 9: Participant-level mean absolute error (%) by Orientation and Position. Horizontal conditions are shown for Positions 0–2 followed by vertical conditions for Positions 0–2. Asterisks denote significant pairwise contrasts within the horizontal condition (, $p < .01$; ***, $p < .001$).**

Means reflected these inferential results. Horizontal error increased from Position 0 ($M = 3.21\%$, $SD = 0.93$) to Position 1 ($M = 6.58\%$, $SD = 2.20$) and Position 2 ($M = 9.41\%$, $SD = 2.44$), whereas vertical error remained comparatively stable (Position 0: $M = 3.39\%$, $SD = 1.37$; Position 1: $M = 3.23\%$, $SD = 1.02$; Position 2: $M = 3.37\%$, $SD = 1.19$). No outliers exceeded ±3SD in any cell.

Figure 10 plots current-study error estimates against the Bezerianos and Isenberg benchmark for descriptive context. All conditions, including the least accurate (Horizontal 2), produced substantially lower absolute error than the historical length benchmark.

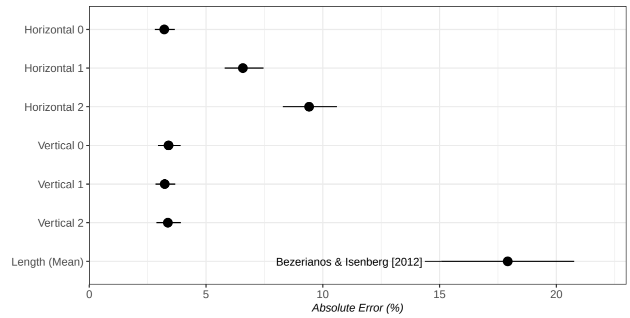

Figure 10: Comparison of absolute error (%) estimates with the Bezerianos and Isenberg [5] historical benchmark for length judgments. Points indicate means and error bars indicate 95% BCa bootstrap confidence intervals. The benchmark value is shown for descriptive context only.

4.2.2 Psychophysical coefficients (individual response curves). We fitted a psychophysical power function to each participant's data for each Orientation × Position combination (see Figure 11), as an exploratory and purely descriptive analysis intended to visually characterize systematic perceptual bias and cross-participant heterogeneity.. The resulting coefficients ("accuracy ratios") describe how perceived ratio scales with actual ratio and values close to 1 indicate near-linear perception, and higher values suggest systematic overestimation. For horizontal bar charts, the mean accuracy ratio was 1.01 ($SD = 0.06$) at Position 0, 1.10 ($SD = 0.11$) at Position 1, and 1.26 ($SD = 0.17$) at Position 2. For vertical bar charts, the corresponding coefficients were 1.04 ($SD = 0.07$), 1.01 ($SD = 0.05$), and 1.05 ($SD = 0.05$) for Positions 0, 1, and 2, respectively.

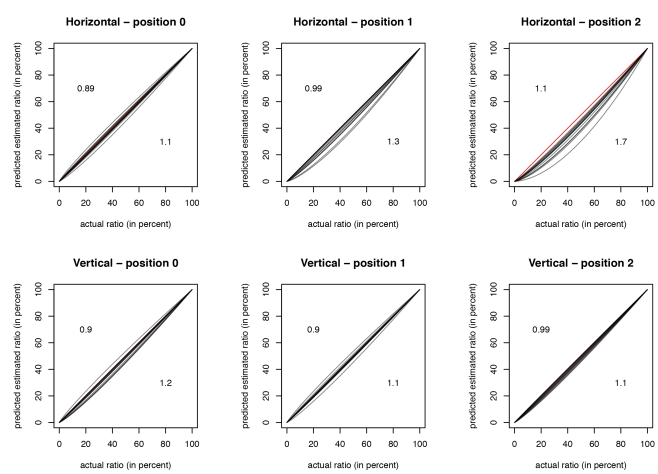

Figure 11: Psychophysical power-function coefficients (accuracy ratios) and individual fitted response curves for Orientation × Position conditions. Horizontal bars show increasing bias and variability, whereas vertical bars remain near linear and stable. The identity line ($y = x$) is shown for reference.

The overplotted individual response curves reflect these patterns. Horizontal bar charts show increasing curvature and between-participant spread as position index increases: functions are close to linear at Position 0, but become progressively steeper and more dispersed at Positions 1 and 2. In contrast, vertical bar charts remain close to the identity line with relatively little variation across positions, indicating near-linear perception and stable accuracy ratios across distances.

5 Discussion

We examined how classic graphical perception findings translate to immersive virtual reality, focusing on visual variables on a virtual table (Experiment 1) and on bar-chart orientation and viewing position on a virtual wall (Experiment 2). Across both experiments, the results generally align with prior 2D graphical perception work while revealing systematic effects of VR-specific viewing conditions.

5.1 Absolute error across visual variables is consistent with prior findings

The reduced error rates observed for bar charts relative to length and area encodings are consistent with Cleveland & McGill's ranking, which places position and length ahead of area and volume for ratio judgments [7, 8]. Our results replicate this ordering in VR: Bar charts yielded the smallest absolute errors, followed by length encodings, and area encodings produced the largest errors. This pattern aligns with 2D studies showing that aligned bars support more accurate judgments than area-based encodings [13, 29, 31, 35]. At the same time, absolute errors in our VR setup were somewhat higher than those typically reported for paper or desktop displays, which is consistent with known distance and scale distortions in immersive environments [9, 20, 26, 38].

The psychophysical coefficients offer a more detailed picture. Bar charts produced coefficients close to one with low between-participant variability, indicating near-linear perception of bar length in VR and reinforcing Cleveland & McGill's claim that position/length is a robust channel [7]. Length encodings showed higher variability and coefficients moderately above one, suggesting modest bias and increased dispersion similar to the "small judgment errors" reported by Spence for length-based displays [31]. Area encodings led to the largest coefficients and variation, which mirrors findings that area judgments are less precise than length judgments on conventional displays [7, 16, 31]. These results support H1 and show that the established ranking of bar, length, and area encodings largely carries over to immersive VR, in line with Watson et al.'s observation that relative percentage judgments can remain accurate in head-mounted displays despite absolute-scale distortions [37].

5.2 Orientation and Distance Interact on Virtual Wall Displays

Experiment 2 extends Bezerianos & Isenberg's findings on large wall displays to VR, where viewing angle and location differentially affect visual variables [5]. We found strong main effects of both orientation and position on absolute error, supporting H2 and H3, as well as a pronounced Orientation × Position interaction. Horizontal bar charts produced higher errors and greater between-participant variability than vertical bar charts (collapsed across positions), and accuracy decreased monotonically from closest to farthest positions (collapsed across orientations). This reinforces prior immersive studies reporting advantages for height-based encodings [1, 25].

The interaction reveals that these effects are not additive: distance primarily harmed horizontal bar charts while vertical bar charts remained robust across viewing positions. At the closest position, horizontal and vertical errors were similar; at medium and far positions, horizontal errors increased sharply while vertical errors stayed stable. Psychophysical coefficients reflect this pattern: horizontal coefficients rose and became more dispersed with distance (1.01 to 1.26), whereas vertical coefficients remained near one (1.04 to 1.05) with modest variability. This selective degradation suggests horizontal bars are more vulnerable to VR distortions at oblique angles [9, 20, 38]. Conversely, our digitized values from Bezerianos & Isenberg indicate their wall errors for length were around 18% at close positions [5]; our VR errors were generally lower, however, vertical orientations yielded high accuracy across all viewing positions, while horizontal bar charts became substantially less accurate as distance and viewing angle increased. Comparatively, Bezerianos & Isenberg found horizontal length judgments to be relatively unaffected by viewing distance.

5.3 Geometric and Optical Mechanisms Underlying Oblique Viewing Costs

Oblique viewing of two-dimensional stimuli foreshortens the retinal projection along the axis perpendicular to the line of sight, with the visual system ordinarily compensating through a correction analogous to shape constancy [11, 23]. On physical media and wall displays, this correction is supported by reliable environmental depth cues such as surface texture gradients and explicit spatial anchors. Bezerianos & Isenberg's tiled display provided the latter in the form of physical bezels between panels, which participants could use as length references [5]. In VR, both categories of support are attenuated: although our virtual wall featured a PBR texture, the rendered environment still provided a relatively sparse and impoverished set of depth cues in comparison with physical viewing without equivalent reference structures. Without these, the perceptual correction for perspective foreshortening is incomplete, and estimation errors accumulate with viewing angle. This accounts for a key asymmetry between our findings and those of Bezerianos & Isenberg: their horizontal bar judgments were relatively unaffected by viewing distance, whereas ours degraded monotonically, an outcome consistent with the removal of bezel-based spatial anchors rather than a fundamental change in the perceptual hierarchy.

VR-specific optical characteristics compound this geometric effect. Single-element HMD lenses introduce transverse chromatic aberration (TCA) that becomes measurable beyond approximately 6° from the centre of the field of view [4], producing colour fringing that degrades the sharpness of bar edges at lateral gaze positions. Vertical bars, whose critical dimension is orthogonal to the axis of foreshortening at any lateral offset, are intrinsically robust to this form of degradation. Horizontal bars, by contrast, have their length dimension aligned precisely with the axis that undergoes both geometric compression and peripheral optical blurring, making accurate length extraction disproportionately harder at distance. Finally, the vergence–accommodation conflict (VAC) inherent to fixed-focus HMD optics introduces systematic depth distortion through a device-specific vergence offset [14], which may amplify scale misjudgements at far positions. In Experiment 1, obliqueness was uniform across all three encoding types, as all stimuli were viewed from the same fixed downward angle to the tabletop, and therefore cannot explain the differential errors between bar, length, and area encodings. It plausibly contributes, however, to the elevated errors relative to paper-based baselines [9, 20].

5.4 Implications and Limitations

Our evidence has straightforward design implications. In immersive analytic scenarios, vertical bar charts placed close to and roughly face-on to the viewer appear to be the safest choice when precise judgments are required. Horizontal bar charts should be used cautiously on wall-like surfaces, particularly at positions that require large head rotations or peripheral viewing. When a virtual wall layout is necessary, placing critical quantitative information at closer or more central positions and using vertical rather than horizontal encodings may help mitigate VR-induced distortions.

Our study has limitations that should be considered. The number of participants is modest, though comparable to many HCI and visualization experiments [6], which constrains the analysis of individual differences. Furthermore, specific demographic information was not available for the present analysis because the original participant questionnaire records were lost following an expiration of access to the data collection platform. The lack of age, gender, and baseline expertise data limits our ability to generalize findings. Specifically, it makes it difficult to isolate whether the magnitude of differences observed in Experiment 2 stems entirely from the VR environment or partially from undocumented demographic variations such as higher-than-average experience with VR and data visualization.

The stimuli were intentionally simple to isolate basic visual variables and positions, which supports internal validity but may reduce generalizability to more complex charts and tasks such as trend or pattern detection [16, 31]. The work also focuses on ratio estimation, which is susceptible to cognitive biases such as rounding, anchoring, and limited numerical precision, and may distort participants' scaling of perceived stimuli [33]. Furthermore, because fine-tuning the exact integer response via the continuous frame-based thumbpad input incurred a time cost, it is likely that participants occasionally 'satisficed' by submitting a sufficiently close value rather than correcting overshoots to hit a highly specific number. Finally, our results are specific to a particular HMD and tracking setup, and hardware characteristics such as field of view or resolution are known to affect depth perception [3, 19, 26]. Additionally, because we restricted our spatial conditions to the horizontal axis to preserve ergonomic comfort and manage session length, our findings do not account for potential vertical obliqueness effects, which warrants future investigation.

Future work could extend these experiments to larger and more diverse samples, more varied chart types and tasks, and direct comparisons between VR, large physical displays, and desktop conditions. Combining our behavioral measures with eye tracking or head movement data in VR, as suggested in recent immersive visualization work [12, 18], may also help clarify how users inspect charts at different positions and orientations and how these behaviors relate to the observed error patterns, as well as elucidate cognitive and physiological states during graphical perception tasks.

6 Conclusion

We presented two experiments examining two-dimensional graphical perception in VR, conceptually replicating established tabletop and wall display paradigms in an immersive setting. Our results show that the classic ranking of visual variables largely holds in VR, with bar charts supporting more accurate judgments than length or area encodings. On virtual wall displays, chart orientation and viewing distance interact in ways that penalize horizontal encodings at peripheral positions while leaving vertical encodings largely unaffected. As immersive environments are increasingly used for information work, these findings provide an empirical basis for design decisions about encoding quantitative data in VR, and indicate that established 2D guidelines cannot always be applied without accounting for the viewing conditions specific to immersive displays.

Data Availability

All analysis scripts and experimental materials are publicly available on OSF[1].

Acknowledgments

The authors acknowledge Dr. Yvonne Jansen for her supervision and guidance during the original project that formed the basis of this article. The authors also thank the support of Prof. Dr. Martin Meißner at the Center for Digital Transformation of the Technical University of Munich. Additionally, this work is supported by the German Research Foundation (DFG), CRC 1404: »FONDA: Foundations of Workflows for Large-Scale Scientific Data Analysis« (Project-ID 414984028).

References

[1] Somnath Arjun, G. S. Rajshekar Reddy, Abhishek Mukhopadhyay, Sanjana Vinod, and Pradipta Biswas. 2021. Evaluating Visual Variables in a Virtual Reality Environment. BCS Learning & Development, 11–22. doi:10.14236/ewic/HCI2021.1
[2] C. Armbrüster, M. Wolter, T. Kuhlen, W. Spijkers, and B. Fimm. 2008. Depth Perception in Virtual Reality: Distance Estimations in Peri- and Extrapersonal Space. *CyberPsychology & Behavior* 11, 1 (Feb. 2008), 9–15. doi:10.1089/cpb.2007.9935
[3] Felipe Bacim, Eric Ragan, Siroberto Scerbo, Nicholas F. Polys, Mehdi Setareh, and Brett D. Jones. 2013. The effects of display fidelity, visual complexity, and task scope on spatial understanding of 3D graphs. In *Proceedings of Graphics Interface 2013 (GI '13)*. Canadian Information Processing Society, CAN, 25–32. https://dl.acm.org/doi/10.5555/2532129.2532135
[4] Ryan Beams, Andrea S. Kim, and Aldo Badano. 2019. Transverse chromatic aberration in virtual reality head-mounted displays. *Optics Express* 27, 18 (Sept. 2019), 24877. doi:10.1364/OE.27.024877
[5] Anastasia Bezerianos and Petra Isenberg. 2012. Perception of Visual Variables on Tiled Wall-Sized Displays for Information Visualization Applications. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Dec. 2012), 2516–2525. doi:10.1109/TVCG.2012.251
[6] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 981–992. doi:10.1145/2858036.2858498
[7] William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (Sept. 1984), 531–554. doi:10.1080/01621459.1984.10478080 _eprint: https://doi.org/10.1080/01621459.1984.10478080.

[1] https://osf.io/45ukz/overview?view_only=2df439172da94636a92463ba08a68bb3

[8] William S. Cleveland and Robert McGill. 1986. An experiment in graphical perception. *International Journal of Man-Machine Studies* 25, 5 (Nov. 1986), 491–500. doi:10.1016/S0020-7373(86)80019-0
[9] James E. Cutting. 1997. How the eye measures reality and virtual reality. *Behavior Research Methods, Instruments, & Computers* 29, 1 (March 1997), 27–36. doi:10.3758/BF03200563
[10] Omar H. Fares, Joseph Aversa, Seung Hwan Lee, and Jenna Jacobson. 2024. Virtual reality: A review and a new framework for integrated adoption. *International Journal of Consumer Studies* 48, 2 (2024), e13040. doi:10.1111/ijcs.13040 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ijcs.13040.
[11] E. Bruce Goldstein. 1987. Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. *Journal of Experimental Psychology: Human Perception and Performance* 13, 2 (1987), 256–266. doi:10.1037/0096-1523.13.2.256
[12] Dongyun Han, Anastasia Bezerianos, Petra Isenberg, and Isaac Cho. 2025. Perception of Visual Variables on Virtual Wall-Sized Tiled Displays in Immersive Environments. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (May 2025), 3045–3055. doi:10.1109/TVCG.2025.3549190
[13] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 203–212. doi:10.1145/1753326.1753357
[14] David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision* 8, 3 (March 2008), 33. doi:10.1167/8.3.33
[15] Yvonne Jansen, Pierre Dragicevic, and Jean-Daniel Fekete. 2013. Evaluating the efficiency of physical visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 2593–2602. doi:10.1145/2470654.2481359
[16] Yvonne Jansen and Kasper Hornbæk. 2016. A Psychophysical Investigation of Size as a Physical Variable. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016), 479–488. doi:10.1109/TVCG.2015.2467951
[17] Yuan Jing, Mansouri Behzad, Pettey Jeff H, Ahmed Sarah Farukhi, and Khaderi S Khizer. 2018. The Visual Effects Associated with Head-Mounted Displays. *International Journal of Ophthalmology and Clinical Research* 5, 2 (June 2018). doi:10.23937/2378-346X/1410085
[18] Matthias Kraus, Karsten Klein, Johannes Fuchs, Daniel Keim, Falk Schreiber, and Michael Sedlmair. 2021. The Value of Immersive Visualization. *IEEE Computer Graphics and Applications* 41, 4 (July 2021), 125–132. doi:10.1109/MCG.2021.3075258
[19] Joseph J. LaViola, Ernst Kruijff, Ryan P. McMahan, Doug Bowman, and Ivan P. Poupyrev. 2017. *3D User Interfaces: Theory and Practice* (2 ed.). Addison-Wesley, Boston.
[20] Jack M. Loomis and Joshua M. Knapp. 2003. Visual Perception of Egocentric Distance in Real and Virtual Environments. In *Virtual and Adaptive Environments*. CRC Press. Num Pages: 26.
[21] Joshua McVeigh-Schultz and Katherine Isbister. 2022. A "beyond being there" for VR meetings: envisioning the future of remote work. *Human–Computer Interaction* 37, 5 (Sept. 2022), 433–453. doi:10.1080/07370024.2021.1994860 _eprint: https://doi.org/10.1080/07370024.2021.1994860.
[22] Gregory B. Newby. 2002. Empirical Study of a 3D Visualization for Information Retrieval Tasks. *Journal of Intelligent Information Systems* 18, 1 (Jan. 2002), 31–53. doi:10.1023/A:1012968507789
[23] David N. Perkins. 1973. Compensating for distortion in viewing pictures obliquely. *Perception & Psychophysics* 14, 1 (Feb. 1973), 13–18. doi:10.3758/BF03198608
[24] Lewis V. Peterson and Wilbur Schramm. 1954. How Accurately Are Different Kinds of Graphs Read? *Audio Visual Communication Review* 2, 3 (1954), 178–189. https://www.jstor.org/stable/30218399
[25] Quang Quach and Bernhard Jenny. 2020. Immersive visualization with bar graphics. *Cartography and Geographic Information Science* 47, 6 (Nov. 2020), 471–480. doi:10.1080/15230406.2020.1771771 _eprint: https://doi.org/10.1080/15230406.2020.1771771.
[26] Rebekka S. Renner, Boris M. Velichkovsky, and Jens R. Helmert. 2013. The perception of egocentric distances in virtual environments - A review. *ACM Comput. Surv.* 46, 2 (Dec. 2013), 23:1–23:40. doi:10.1145/2543581.2543590
[27] Pericles 'asher' Rospigliosi. 2022. Metaverse or Simulacra? Roblox, Minecraft, Meta and the turn to virtual reality for education, socialisation and work. *Interactive Learning Environments* 30, 1 (Jan. 2022), 1–3. doi:10.1080/10494820.2022.2022899 _eprint: https://doi.org/10.1080/10494820.2022.2022899.
[28] Marc M. Sebrechts, John V. Cugini, Sharon J. Laskowski, Joanna Vasilakis, and Michael S. Miller. 1999. Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '99)*. Association for Computing Machinery, New York, NY, USA, 3–10. doi:10.1145/312624.312634
[29] David Simkin and Reid Hastie. 1987. An Information-Processing Analysis of Graph Perception. *J. Amer. Statist. Assoc.* 82, 398 (June 1987), 454–465. doi:10.1080/01621459.1987.10478448
[30] Alexis D. Souchet, Domitile Lourdeaux, Alain Pagani, and Lisa Rebenitsch. 2023. A narrative review of immersive virtual reality's ergonomics and risks at the workplace: cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload. *Virtual Reality* 27, 1 (March 2023), 19–50. doi:10.1007/s10055-022-00672-0
[31] Ian Spence. 1990. Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance* 16, 4 (1990), 683–692. doi:10.1037/0096-1523.16.4.683
[32] S. S. Stevens. 1957. On the psychophysical law. *Psychological Review* 64, 3 (1957), 153–181. doi:10.1037/h0046162
[33] S. S. Stevens. 1971. Issues in psychophysical measurement. *Psychological Review* 78, 5 (1971), 426–450. doi:10.1037/h0031324
[34] Christian M. Stracke, Pia Bothe, Simon Adler, Evelyn Susanne Heller, Jonas Deuchler, Jenny Pomino, and Matthias Wölfel. 2025. Immersive virtual reality in higher education: a systematic review of the scientific literature. *Virtual Reality* 29, 2 (April 2025), 64. doi:10.1007/s10055-025-01136-x
[35] Justin Talbot, Vidya Setlur, and Anushka Anand. 2014. Four Experiments on the Perception of Bar Charts. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 2152–2160. doi:10.1109/TVCG.2014.2346320
[36] James R. Wallace, Daniel Vogel, and Edward Lank. 2014. The effect of interior bezel presence and width on magnitude judgement. In *Proceedings of Graphics Interface 2014 (GI '14)*. Canadian Information Processing Society, CAN, 175–182. https://dl.acm.org/doi/10.5555/2619648.2619678
[37] Diane Watson, George Fitzmaurice, and Justin Matejka. 2021. How Tall is that Bar Chart? Virtual Reality, Distance Compression and Visualizations. In *Graphics Interface 2021*. Canadian Human–Computer Communications Society.
[38] Peter Willemsen, Mark B. Colton, Sarah H. Creem-Regehr, and William B. Thompson. 2004. The effects of head-mounted display mechanics on distance judgments in virtual environments. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization (APGV '04)*. Association for Computing Machinery, New York, NY, USA, 35–38. doi:10.1145/1012551.1012558
[39] Jeff Zacks, Ellen Levy, Barbara Tversky, and Diane J. Schiano. 1998. Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology: Applied* 4, 2 (1998), 119–138. doi:10.1037/1076-898X.4.2.119

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009