Revisiting Channel Effectiveness: A Multi-Dimensional Evaluation with Primitive Visual Stimuli

Soohyun Lee, Seokhyeon Park, Minsuk Chang, and Jinwook Seo

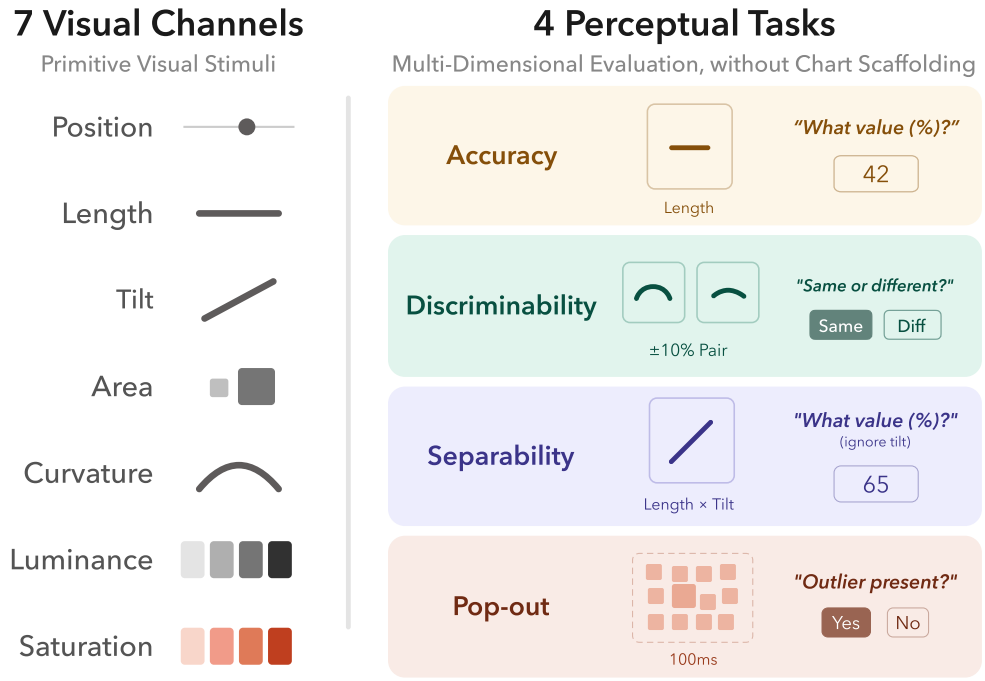

Fig. 1: We evaluate seven visual channels (left) across four perceptual tasks: accuracy, discriminability, separability, and pop-out. Every task uses primitive stimuli stripped of chart scaffolding. Each row shows an example stimulus, the question posed to participants, and a sample response (right).

Abstract—Established channel effectiveness rankings primarily assess magnitude estimation accuracy in complete chart contexts, often neglecting other perceptual tasks such as discriminability, separability, and pop-out. To address this gap, we conducted crowdsourced experiments on seven core visual channels (position, length, tilt, area, curvature, luminance, and saturation) using primitive visual stimuli, a set of visual marks without chart-specific scaffolding to isolate channel-level variation. We evaluated these channels across four perceptual tasks (accuracy, discriminability, separability, and pop-out) and found that channel effectiveness is fundamentally multi-dimensional, with rankings shifting substantially across tasks. For instance, while spatial channels maintain an overall advantage, accuracy depends strongly on whether a fixed spatial anchor is available. Discriminability varies dramatically across channels and value ranges, a pattern we formalized with a novel Anchored Harmonic Weber model. Pairwise channel interactions are often strongly asymmetric. Finally, we identify a dissociation between estimation accuracy and preattentive detection: length shows only moderate detection effectiveness despite top-tier accuracy, while area achieves the highest detection rates despite poor quantitative accuracy, though the latter advantage may partly reflect stimulus-level cues. We synthesize these findings into a scenario-driven perspective for context-sensitive channel selection.

Index Terms—Visual channels, graphical perception, channel effectiveness, primitive visual stimuli

1 Introduction

Visual channels such as position, length, tilt, area, and color are the fundamental building blocks of data visualization. Cleveland and McGill's foundational work [13], later corroborated by Heer and Bostock [30], established influential rankings showing that spatial channels (position, length) enable more accurate reading than others (angle, area, color). This hierarchy has shaped design guidelines and automated visualization systems [39] for decades.

However, this knowledge faces critical limitations in modern visualization design. First, classic experiments tested channels within complete chart frameworks (bar charts with axes, scatter plots with gridlines), making it difficult to separate channel behavior from the perceptual scaffolding these chart elements provide [13]. For instance, the perceptual accuracy of a pie chart varies depending on whether viewers rely on angle versus area [32,46], and radial layouts produce conflicting outcomes depending on the channel used [56]. Furthermore, existing

• *Soohyun Lee, Seokhyeon Park, and Jinwook Seo are with Seoul National University. E-mail: {shlee, shpark}@hcil.snu.ac.kr; jseo@snu.ac.kr.*
• *Minsuk Chang is with Georgia Institute of Technology. E-mail: minsuk@gatech.edu.*
• *Jinwook Seo is the corresponding author.*

Manuscript received xx xxx. 202x; accepted xx xxx. 202x. Date of Publication xx xxx. 202x; date of current version xx xxx. 202x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.202x.xxxxxxx

work focuses primarily on accuracy while neglecting other essential perceptual tasks: discriminability (detecting small differences), separability (combining multiple channels), and pop-out (rapidly identifying outliers) [23, 42].

We address these gaps by testing seven channels (position, length, tilt, area, curvature, luminance, saturation) across four tasks (accuracy, discriminability, separability, and pop-out). To do so, we evaluate channels using primitive visual stimuli, visual marks stripped of chart-specific scaffolding (e.g., a single line segment within a shared display frame), to isolate channel-level variation.

Our results both confirm and challenge established knowledge on channel effectiveness. Spatial channels retain their accuracy advantage, consistent with classic rankings, but cross-task profiles reveal striking dissociations, most notably, area performs poorly in accuracy yet achieves the highest detection rate in our pop-out setting. Such reversals show that channels sidelined by accuracy-centric rankings play a central role once performance is evaluated across multiple tasks.

Building on this multi-dimensional view, we use an equivalence margin and effect sizes to distinguish ranking differences that are practically meaningful from those that are negligible (for instance, tilt's accuracy is statistically equivalent to single position). We synthesize these results into a scenario-driven perspective that maps each task to a recurring design dimension, enabling context-sensitive encoding choices beyond traditional chart conventions.

Our main contributions are:

1. A systematic evaluation of seven visual channels across four perceptual tasks (accuracy, discriminability, separability, and pop-out), with channel subsets and pairings adapted per task, using primitive visual stimuli that isolate channel-level variation from chart-specific scaffolding.
2. Structured evidence that channel effectiveness is multi-dimensional: channel rankings shift substantially across the four tasks, and pairwise channel interactions exhibit pronounced asymmetries, demonstrating that no single ordering captures perceptual performance comprehensively.
3. A scenario-driven perspective that organizes these findings into four recurring design dimensions (task stakes, data granularity, background interference, and response time) to guide context-sensitive encoding choices.
4. Evidence of a dissociation between estimation accuracy and preattentive detection: length shows only moderate pop-out despite top-tier accuracy, while area achieves the highest detection rate despite poor quantitative accuracy (though area's advantage may partly reflect stimulus-level cues).

2 Related Work

2.1 Graphical Perception in Visualization Contexts

Cleveland and McGill [13] compared bar, pie, and scatter plots and showed that position and length enable lower error than angle or area in complete charts, and Mackinlay [39] codified this ranking into automated design heuristics, tying perceptual accuracy to chart form rather than to the underlying channel. These channel taxonomies trace back to Bertin's visual variables [5]. Heer and Bostock [30] confirmed these findings via crowdsourcing, yet their stimuli retained extensive scaffolding (e.g., axes, tick marks, gridlines, and legends). Such scaffolding can inflate or depress accuracy. Axes simplify height estimation, while adjacent bubbles may distract or artificially simplify the task. More recent computational work has applied CNNs to chart-level graphical perception [26], probed the sensitivity of vision models to isolated channels [36], and disentangled visual reading from factual priors in LVLMs' visualization literacy [37], but none of these studies combined human observers with context-free, single-channel stimuli.

More broadly, recent work has questioned whether a single channel ranking can serve all visualization tasks, showing that channel effectiveness depends on perceptual task and data distribution [33, 40, 44]. These findings motivate our multi-task evaluation rather than relying on a single performance dimension.

2.2 Perceptual Foundations of Psychophysics and Channel Interactions

Psychophysical theory links physical and perceived magnitude. *Stevens' power law* predicts near-linear growth for length but compressive curves for brightness and area [49], whereas the *Weber–Fechner* relation formalizes JND thresholds that underpin color-scale design [20].

Subsequent work highlights categorical anchors. Orientation adaptation studies show that 0°, 45°, and 90° act as attractors, biasing estimates toward canonical angles [10, 17, 57]. In color perception, Szafir derived difference functions for luminance and hue that inform palette design [51], and Bartram et al. demonstrated that grid-line color together with transparency jointly affect detection thresholds for overlaid grids [2]. These efforts reveal that perception is highly non-linear, even for single channels. Yet systematic validation across the palette of encodings, particularly saturation and tilt, remains limited.

Beyond single-channel perception, preattentive processing (rapid, parallel detection of salient features before focused attention) is central to visualization design. Feature Integration Theory [54] established that basic features are registered in parallel while conjunctions require serial search [59], and Healey and Enns identified which visual properties support rapid detection in visualization contexts [29]. Yet most pop-out studies in the visualization literature test features embedded in multi-element displays, leaving open whether the same preattentive hierarchies hold when chart-level scaffolding is absent.

When multiple channels are combined, the question of independence arises. Garner's separable–integral framework [24] distinguishes channels that can be attended independently from those that interfere during perceptual judgment. Smart and Szafir [47] measured the separability of shape, size, and color in scatterplots, finding significant interactions that violate simple independence assumptions. However, systematic pairwise separability testing across a broader channel set remains lacking, particularly for geometric channels such as tilt and curvature, which have received limited empirical attention in visualization. Our study addresses these gaps by unifying power-law fitting, Weber-fraction analysis, pairwise separability testing, and preattentive detection within a single crowdsourced protocol.

2.3 Primitive-Stimulus and Single-Channel Experiments

Recently, researchers have begun stripping chart context from the stimuli to conduct more precisely controlled experiments. Szafir fit JND curves from single color swatches against a background [51], Veras and Collins manipulated exactly one glyph attribute per trial to compare human similarity ratings against an image-based discriminability metric [55], and Demiralp et al. learned perceptual kernels from crowdsourced judgments over visualization stimuli [16]. Bartram et al. removed data marks entirely, varying only grid-line color and transparency to establish detection thresholds [2], while Bearfield et al. used unlabeled two- or three-bar arrays and uncovered length-comparison biases even under this reduced context [3].

While each of these studies reduces stimuli, most retain contextual or competing elements, such as multiple glyphs, background grids, or adjacent bars, that can redirect attention. They also tend to focus on a single channel and report either magnitude error or threshold sensitivity, but rarely both, leaving channels such as saturation and tilt untested under fully isolated conditions.

We extend this paradigm. A single mark varies along one channel with all other attributes fixed, letting us measure magnitude-estimation accuracy and fine-grained discriminability from the same observers.

3 Method

We conducted crowdsourced experiments with primitive visual stimuli to characterize channel behavior in a controlled setting that removes chart-specific scaffolding while retaining a shared display frame, testing accuracy (estimation precision), discriminability (just-noticeable differences), separability (cross-channel interference), and pop-out across seven visual channels. We refer to these four conditions as tasks because each corresponds to a distinct perceptual operation that arises in real-world chart reading. Figure 1 provides an overview of the experimental framework and task structure across these four tasks.

3.1 Stimuli Design

Each stimulus consisted of exactly one visual mark (e.g., a line segment or filled square) that varied along a single encoding channel, while all other attributes remained fixed at default values. All stimuli were displayed on a white background within a square canvas of 500×500 px bounded by a 1 px black border. The canvas provides a shared display frame and bounds the drawable extent of the length and area channels; position is referenced to its line segment, while the color and angular ranges are established by the pre-block reference below. We use *primitive* to denote deconstructed marks that remove chart-specific scaffolding (axes, gridlines, labels, legends, adjacent marks) to isolate channel-level variation, retaining only the canvas boundary as a shared frame. Isolating individual visual channels for controlled perceptual measurement has precedent in visualization research [6, 36]. Before each experiment block, participants viewed a reference display showing the full range of the target channel (e.g., a color bar spanning 0% to 100% for luminance and saturation), ensuring comparable task references across spatial and chromatic channels. Because every channel is judged relative to the on-screen canvas or the pre-block reference rather than in absolute physical units, we did not perform per-display calibration, as is standard for crowdsourced studies [30]. Variation in display hardware and viewing conditions therefore remains a limitation, particularly for the color channels.

We examined seven channels: position, length, tilt, area, curvature, luminance, and saturation, spanning the major magnitude channels emphasized in visualization effectiveness frameworks such as Munzner's [42]. Position was further subdivided into three reference-frame variants detailed in Section 4. Hue and shape were included only in the pop-out experiment because the other three tasks require ordinal magnitude estimation on a bounded linear scale, which they lack without imposing an arbitrary origin (see Section 7).

Metric channels (length, position, area, luminance, saturation) used integer steps from 0 to 100, while tilt and curvature used 0°–180° in 1° increments. For area, stimuli were filled squares spanning the smallest to largest shape allowed within the display region. For curvature, the two endpoints matched the length stimulus, while the connecting path formed a circular arc whose central angle ranged from 0° (straight) to 180° (semicircle).

Default stimulus parameters (single centered line; 50% length, 0° tilt, 50% luminance, 0% saturation, 0° curvature, red hue (0°)) provided a neutral baseline so that non-target channels would not distract from the channel under investigation. The color stimuli fix the HSL hue at the default red (0°); luminance follows the HSL lightness axis (0% = black to 100% = white) and saturation the HSL saturation axis (0% = gray to 100% = fully saturated red), each mapping its 0–100% value directly onto that coordinate.

3.2 Participants and Procedure

All experiments were implemented and deployed using the reVISit framework [14, 18][1], an open-source platform for conducting scalable online visualization studies. Each stimulus was implemented as a React component within the reVISit environment, enabling programmatic generation of primitive visual stimuli with precise control over channel parameters. We leveraged reVISit's built-in support for trial randomization, response capture, and integration with Amazon Mechanical Turk (MTurk) for participant recruitment and routing. Our study was within-subjects, and we recruited 119 participants through MTurk, limiting eligibility to workers with at least a 97% approval rate and 5000 or more approved HITs. Across the core participant pool, 58% identified as male and 42% as female, the mean age was 36.9 years ($\pm$12.1 SD), and 99% reported at least some college education. Attention-check trials presented stimuli with unambiguous ground-truth values (e.g., clearly extreme values such as 0 or 100 for metric channels, or maximally distinct stimulus pairs for discriminability trials); 14 participants who responded incorrectly on these trials were excluded from all analyses. All 105 core participants completed all four tasks (Accuracy, Separability, Discriminability, and Pop-out), and we recruited an additional 45 participants for Discriminability to obtain more fine-grained samples.

The study took approximately 12 minutes in total, where participants provided informed consent, completed the assigned task blocks, and reported demographics (age range, gender, and education level). Before the main task blocks, participants completed one or two practice rounds per task to ensure they understood the response formats. Both task-block order and channel presentation order were fully randomized per participant to mitigate learning and fatigue effects. For each main trial, participants typed numeric values (Accuracy and Separability), chose "Same/Different" (Discriminability), or indicated outlier presence (Pop-out). Trials were self-paced with no explicit time limit; the pop-out task additionally enforced a 100 ms stimulus exposure in a guided, uniform-distractor design, as described in Section 7. Trial counts varied by task: Accuracy (2), Discriminability (10 for the core participants; 32 for the additional participants), Pop-out (3), and all tested primary-secondary combinations in Separability; Accuracy and Pop-out counts were kept low to prioritize breadth across 7 channels $\times$ 4 tasks, whereas Discriminability required denser sampling to estimate JND contours. Participants were compensated $2 upon completion of all assigned trials, approximately $10 per hour. The Seoul National University IRB approved the protocol (No. 2503/004-016). All four task blocks can be run at https://graphical-perception.github.io/Primitive-Study/.

4 Accuracy

4.1 Task Design

In each trial, participants performed a magnitude-estimation task, observing a single primitive stimulus and estimating the target channel's magnitude on its natural scale (0–100% for metric channels; 0°–180° for tilt and curvature).

We tested seven channels: position, length, tilt, area, luminance, saturation, and curvature. For the **position** channel, we designed three variants to probe how reference frames affect magnitude judgment:

- *Single position:* A single line segment is presented with a marked point; participants estimate what percentage from the start endpoint the mark appears. Both horizontal and vertical reference lines were used (start endpoint at the left or bottom, respectively). Note that this task is functionally equivalent to judging length from a fixed anchor.
- *Position comparison (aligned):* Two identical, parallel line segments are displayed, each with a marked point, sharing a common baseline (their start endpoints lie on the same line). Participants report the absolute difference (in percentage points) between the relative positions of the two marked points along their segments.
- *Position comparison (unaligned):* The same two parallel segments are presented, but each is translated in the opposite direction along its own length axis, so their start endpoints no longer share a common baseline. Participants again estimate the absolute difference (in percentage points) between the relative positions of the marked points.

Stimulus images for all variants and channels appear in Figure S1. For the **length** channel, participants judged a center-aligned horizontal line segment that extended bidirectionally from its midpoint, estimating its length as a percentage of the maximum possible length (0–100%).

We obtained data from 105 participants, with two responses per participant for each channel condition (210 responses per channel). For each trial, a value (or pair of values) was randomly selected within the channel's domain, with all true percentages uniformly distributed across the possible range.

4.2 Analysis

To analyze perceptual accuracy, we compared each participant's response (P) to the ground-truth value (S), both normalized to $[0,1]$ by the channel maximum $S_{\max}$ (e.g., 100 for length, 180 for tilt). We modeled the response as a power-law transform S^α, with α either fixed at 1 (baseline linear model) or fit per channel by a fine grid search over $\alpha \in [0.25, 4]$ minimizing the mean log-error. The fitted exponents appear in Figure 2.

[1] https://revisit.dev/

$$\text{Mean log-error} = \frac{1}{N}\sum_{i=1}^{N}\log_2\left(|P_i - S_i^{\alpha}| + \tfrac{1}{8}\right) \quad (1)$$

The additive $\frac{1}{8}$ offset, adapted from Cleveland and McGill [13], prevents the logarithm from diverging at zero error; its effect is examined in Supplementary S3.1. All pairwise comparisons used paired t-tests on per-participant means, with Benjamini–Hochberg FDR correction (FDR level 0.05, 36 comparisons across nine conditions and seven channels, with position subdivided into three reference-frame variants). We chose Benjamini–Hochberg over Bonferroni because the large number of comparisons makes family-wise error control overly conservative, whereas BH maintains statistical power while controlling the expected proportion of false discoveries [4]. We checked the normality of the paired differences with Shapiro–Wilk tests and cross-checked headline comparisons with Wilcoxon signed-rank tests (Supplementary S3.2). For channels hypothesized to perform equivalently, we additionally applied two one-sided tests (TOST) to assess whether observed differences fell within a pre-specified equivalence margin.

4.3 Results

Figure 2 summarizes the mean log-error for each channel, comparing the baseline linear model (blue) and best-fit power-law model (orange). Lower values indicate higher perceptual accuracy. The reported comparisons below survived FDR correction. Cross-channel comparisons are shaped in part by the display frame, which offers stronger within-trial reference cues to spatial channels than to color (Section 8).

Position (single) achieved the highest accuracy, significantly outperforming **length** (paired t-test, $p = 0.009$, Hedges' $g = 0.292$, FDR-corrected). Unlike single position, which is read from a fixed end anchor, the length stimulus extends bidirectionally from its midpoint. This bidirectional extension produced systematic overestimation ($\alpha = 0.759$), particularly at smaller values, consistent with the cognitive cost of integrating two segments from a center point.

Position (aligned) condition, where participants compared positions on two horizontally aligned lines, performed worse than both the single position and center-aligned length. **Position (unaligned)** condition, whose segments lack a shared start-point reference, showed further degradation. This progression from single position to unaligned comparison demonstrates how perceptual accuracy deteriorates systematically as reference frames become less stable.

Area exhibits nearly the poorest baseline accuracy among all channels, yet power-law correction with $\alpha = 0.424$ (95% bootstrap CI $[0.393, 0.490]$; 1,000 participant-level resamples, percentile method) yields the largest improvement across all channels (0.531 log-units), substantially narrowing the gap with position and length.

Color channels show negligible improvement under power-law fitting: **saturation** and **luminance** improved by only 0.042 and 0.032, respectively, an order of magnitude smaller than area's 0.531-unit improvement, indicating more fundamental limitations rather than correctable nonlinearity.

Tilt fell within a pre-specified ± 0.2 log-error equivalence bound relative to single position (TOST: $p = 0.015$; 90% CI $[-0.162, 0.071]$; $d_z = -0.062$), while **curvature** performed comparably to the aligned-position condition. Both show near-linear perception with minimal benefit from power-law fitting.

4.4 Discussion

Reference frame stability shapes spatial accuracy. The single-position advantage over center-extended length (the $\alpha = 0.759$ overestimation noted above) reflects the cognitive cost of bidirectional estimation, and explains why unidirectional measurement from a stable reference point, as in bar charts, affords a perceptual advantage. Even perfectly aligned baselines introduce cognitive load relative to single-anchor judgments, with misalignment compounding the penalty. Because these effects persist with primitive stimuli stripped of axes and gridlines, the perceptual benefits of spatial alignment reflect intrinsic perceptual constraints rather than chart-decoration artifacts.

Power-law correction for the area channel. Area's substantial improvement under power-law correction aligns with Stevens' square-root hypothesis [49], and with classic cartographic work on graduated symbol scaling [21]. This pattern is consistent with the theory that observers estimate area by mentally assessing edge lengths [48] rather than directly perceiving two-dimensional extent. Our design does not separate genuine area compression from a side-length heuristic, since both predict the observed compression, and participants received no specific training to discourage judging side length rather than area. This correction, the largest across all channels, suggests that area's poor reputation largely reflects structured, correctable nonlinearity rather than irreducible noise. The fitted exponent ($\alpha = 0.424$) is close to, though somewhat below, the square-root value (0.5) implied by a side-length heuristic, although reported exponents vary with stimulus geometry (e.g., circles vs. squares) and paradigm. It thus offers a practical calibration value for area encodings.

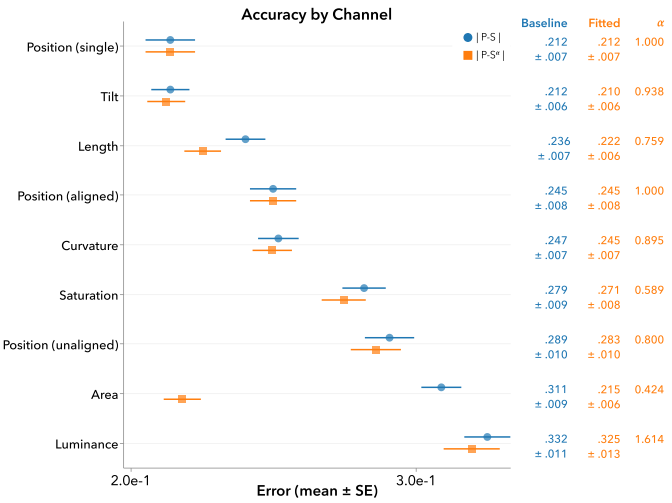

Fig. 2: Per-channel accuracy: baseline (blue, α=1) and power-law corrected (orange, optimal α). Markers show 2^m, the back-transformed mean log-error, with horizontal bars denoting ± 1 SE; lower is more accurate. Values annotated at right.

Tilt achieves nearly the highest accuracy in primitive stimuli. Tilt's accuracy is statistically equivalent to single position (TOST above), contrary to prior chart-level studies [13, 30] that ranked angle-based judgments below position and length. We set the equivalence margin at ± 0.2 log-error because it is directly interpretable on the task's scale and is practically small relative to the larger cross-channel gaps we emphasize. This discrepancy suggests that chart-level gaps arise largely from competing cues, such as area and arc length in pie charts, that obscure the orientation signal, not from weak angular perception. When these confounds are removed, tilt's precision within this equivalence margin rivals that of spatial position (Section 8), suggesting that chart-level angle underperformance is largely contextual rather than intrinsic to angular perception.

5 Discriminability

5.1 Task Design

The discriminability task measures the minimum difference required for reliable visual discrimination, known as the just-noticeable difference (JND) across each channel's range. This captures how sensitive perception is to small changes, which is critical for distinguishing between similar data values in dense visualizations.

In each trial, two stimuli appeared side by side. The left stimulus displayed a randomly selected value within the target channel's domain, while the right stimulus varied by $\pm 10\%$ of that reference value, selected uniformly in $[-10\%, +10\%]$. This $\pm 10\%$ range spans up to five times the Weber fraction for the tested channels (2–8% for spatial and color properties [28, 53]), ensuring each trial pair includes stimuli from near-threshold to clearly suprathreshold.

We tested six channels: length, tilt, area, luminance, saturation, and curvature. Position was excluded because, in this relative same/different

task, position and length reduce to the same judgment of spatial extent; the reference-frame advantage of position (Section 4) applies to absolute estimation, not relative discrimination (see Limitations).

Participants judged whether the two stimuli were "Same" or "Different". A total of 150 participants contributed to the Discriminability experiment. The 105 core participants each completed 10 trials per channel, and an additional 45 participants each completed 32 trials per channel. This unequal allocation was part of the experimental design, improving coverage of the two-dimensional reference-value × offset space for contour extraction and aggregated psychometric fitting to estimate JNDs. All responses were pooled at the trial level before computing response-probability grids, so each trial contributed equally regardless of which participant group generated it; robustness checks for this pooling are reported in Supplementary S3.3.

5.2 Analysis

To quantify how discriminability varies across each channel's domain, we computed the proportion of "Same" responses for each combination of reference value (x) and absolute difference ($|\Delta|$), forming a response probability grid. Contours at levels 10%, 20%, 30%, 40%, and 50% (for the "Same" response rate) were extracted and visualized, providing a compact summary of the response surface. Standard Weber and symmetric boundary models do not capture the asymmetric anchoring patterns evident in these contours, consistent with prior evidence that simple Weber-style models can be insufficient for visualization judgments [31]. We therefore developed the **Anchored Harmonic Weber model**, which extends the classical Weber framework with asymmetric boundary-proximity weighting and fits that shared structure across contour levels. The minimal detectable difference at a given reference value x is modeled as:

$$|\Delta|(x) = w_0 \cdot \frac{x}{x_{max}} + 1/\left(\frac{1}{w_L x} + \frac{1}{w_R(x_{max} - x)}\right) + \text{offset}_i \qquad (2)$$

The first term, $w_0 \cdot \frac{x}{x_{max}}$, captures the baseline Weber fraction. The second term is a harmonic-mean combination of two sensitivity terms anchored at the minimum (w_L, scaling with x) and maximum (w_R, scaling with $x_{max} - x$) ends of the domain. Our model ensures proximity-based dominance consistent with the categorical perception of orientations [10], in which the nearer anchor contributes disproportionately to the threshold. The boundary term vanishes at each endpoint and peaks in the interior, capturing sharper discrimination near either anchor.

Each contour level is allowed an independent vertical offset, but all contours share the same w_L, w_R, and w_0 parameters, ensuring a common underlying shape across contour levels. For the tilt channel, response contours revealed a qualitative regime change at 90° (vertical), where the contour pattern mirrors around the cardinal axis; we therefore fit separate models for $x < 90°$ and $x \geq 90°$. No other channel exhibited a comparable mid-range discontinuity, so all remaining channels were fit with a single model spanning the full domain.

The endpoint slopes report the rate at which thresholds decline toward each boundary (left: $w_L + w_0/x_{max}$ and right: $w_R - w_0/x_{max}$), decomposing each boundary's sensitivity into anchor-driven (w_L, w_R) and baseline Weber (w_0) components.

To confirm that this complexity is warranted, we compared our model against three nested alternatives: constant-JND, standard Weber, and symmetric-anchored models. Our model achieves the lowest BIC across all seven channel conditions ($\Delta\text{BIC} \geq 52$; Supplementary Table S1), confirming that asymmetric boundary anchoring captures structure in the discriminability contours beyond simpler accounts.

5.3 Results

Figure 3 shows the fitted discriminability curves for each channel, estimated from participants' "Same/Different" responses across five JND levels. Channels exhibited diverse discriminability patterns, differing primarily in three respects: (1) how strongly they follow Weber behavior, (2) the number of anchor-like regions shaping the peaks, and (3) boundary effects, where the curve converges to zero toward the endpoints. Where this convergence is strongest (length and area near their empty end, luminance near its bright end), even the smallest differences we tested were detected by a majority of participants, pushing the fitted 40–50% contours below the tested range (Supplementary S1). Fit quality and boundary slopes are summarized in Table 1.

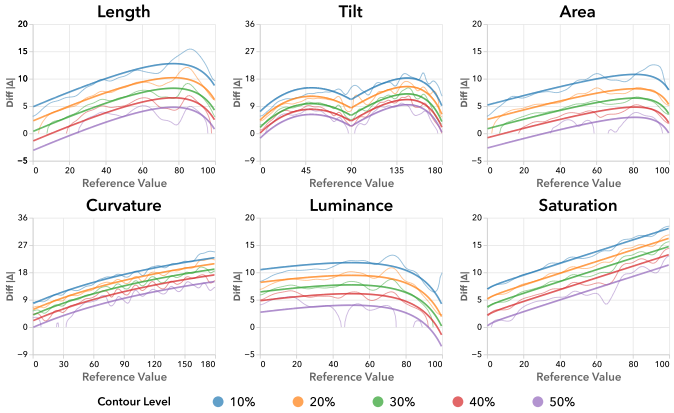

Fig. 3: Discriminability contour maps for each visual channel. The horizontal axis shows the reference value and the vertical axis the minimal detectable difference ($|\Delta|$). For each "Same"-response level (10%, 20%, 30%, 40%, 50%, each a distinct color), the thin line is the contour extracted from the response-probability grid and the thick line is the fitted Anchored Harmonic Weber model. A thin line breaks where the grid has no crossing at a given reference value.

Table 1: Anchored Harmonic Weber model fit (R^2) and boundary slopes for each channel.

Channel	R^2	Left slope	Right slope
Area	0.93	0.0920	0.4175
Curvature	0.96	0.1898	−0.0615
Length	0.89	0.1225	2.0544
Luminance	0.93	0.0545	1.4228
Saturation	0.99	0.1279	−0.0990
Tilt(0°–90°)	0.83	0.1204	1.0296
Tilt(90°–180°)	0.87	0.1490	0.9858

Saturation best approximates Weber's law pattern in our dataset, with JND increasing linearly across the entire range. Unlike all other channels, saturation's thresholds grow proportionally through the full range; the negative right slope confirms the absence of boundary anchoring, indicating that maximum saturation does not function as a perceptual reference. This near-perfect Weber adherence is consistent with ratio-based sensitivity without strong boundary anchors, making saturation the most uniform channel in our discriminability dataset.

Luminance departs substantially from Weber's law. JND increases only marginally across most of the range, indicating a near-constant noise floor, then falls steeply near maximum luminance. This pattern is consistent with a fixed threshold dominating the darker range, while boundary-related sensitivity improves discrimination only near white.

Tilt exhibits the most complex pattern, with three distinct high-sensitivity regions creating a segmented discriminability landscape. We fit separate models for 0°–90° and 90°–180°. The high-sensitivity regions at 0° (horizontal), 90° (vertical), and 180° (horizontal again) create sharp convergence points where discriminability improves dramatically, consistent with prior work on cardinal orientations.

Area shows dual anchoring with remarkably asymmetric slopes. JND decreases toward both empty (0%) and full (100%), but the approach to empty is gradual while the approach to full is steep. This asymmetry indicates that maximum area functions as a stronger boundary reference than empty space, as evidenced by the higher right slope.

Length displays the strongest right-side anchoring in our dataset, with dramatic improvement in discriminability as length approaches maximum. This extreme right-anchoring suggests that the canvas

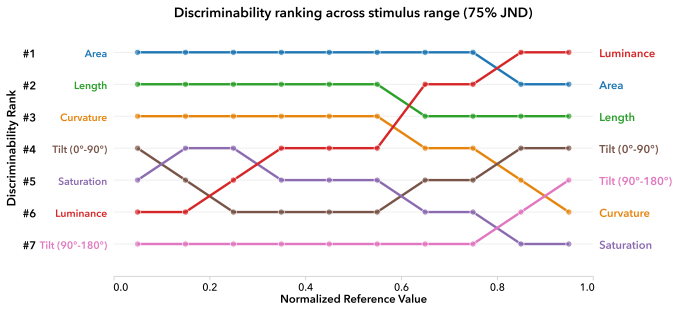

Fig. 4: Discriminability ranking across the stimulus range at the 75% JND threshold. Each line tracks a channel's rank (1 = finest discriminability) as the normalized reference value increases. Rankings shift substantially: area and length dominate at low values, but luminance rises from the bottom ranks to rank 1 near the upper boundary, overtaking all other channels.

boundary provides a strong reference for length judgments.

Curvature follows Weber's law through most of its range but departs from it near 180°. As the arc approaches a semicircle, the discriminability curve bends sharply upward, then plateaus, suggesting a perceptual boundary near the transition from "curved line" to "half circle."

The 25% "Same"-response contour (i.e., 75% of responses indicated "Different"), extracted in the same manner, provides a standardized comparison across channels. Figure 4 tracks each channel's discriminability rank at this threshold across the stimulus range. At low reference values, area and length achieve the finest discriminability, with luminance and tilt ranking lowest. This hierarchy then shifts substantially: luminance rises from the bottom ranks to rank 1 near the upper domain boundary, overtaking all other channels, while curvature and saturation degrade steadily. This dynamic reordering demonstrates that no single discriminability ranking characterizes channel performance across the full stimulus range.

5.4 Discussion

Luminance and saturation occupy opposite extremes of discriminability regularity. Saturation's near-perfect Weber behavior reflects ratio-based processing without strong perceptual references, whereas luminance's flat profile across the darker range suggests absolute-threshold-like processing, where fixed detection limits override proportional scaling [7, 38]. For design, saturation encodings can span the full range with predictable resolution, whereas luminance encodings should favor the bright end where discriminability is sharpest.

Tilt's discriminability reflects a segmented perceptual landscape. Tilt yields the lowest model fit among all channels ($R^2 = 0.83/0.87$), reflecting not measurement noise but a discriminability landscape more complex than a single model can capture. The cardinal orientations (0°, 90°, 180°) act as high-sensitivity anchors, necessitating separate half-range models. Within each half-range, midrange variability remains elevated, consistent with the oblique effect, in which discrimination is poorer at oblique orientations than at cardinal ones [10]. The two half-ranges are not mirror images: discrimination sharpens more steeply toward 90° (from below) and 180° than toward the opposite ends, a left–right asymmetry we detail in Supplementary S1. This segmented landscape coexists with tilt's position-level accuracy (Section 4), demonstrating that average magnitude estimation and local sensitivity patterns are independent aspects of channel performance.

Maximum extent provides a strong perceptual boundary. By *boundary anchoring* we mean that a domain endpoint acts as a perceptual reference, so discrimination sharpens (the contour converges toward zero) as values approach it. Both length and area exhibit pronounced right-skew in their anchoring profiles. Length's right slope (2.054) far exceeds its left slope (0.123), and area shows a comparable pattern (right: 0.418 vs. left: 0.092). This pattern suggests that maximum extent or fullness functions as a stronger perceptual boundary than the corresponding near-empty state. The canvas edge anchors length judgments, and the fully filled region anchors area judgments. Combined with the accuracy finding that a fixed anchor improves magnitude estimation (Section 4), both results indicate that explicit boundary references enhance perception across multiple tasks (Section 8).

Table 2: Channel pairs included in the separability test. A check mark (✓) indicates a tested pair, with the row channel as primary and the column channel as secondary. Blank cells were excluded either because the combination was geometrically infeasible (e.g., area–curvature, area–length conditions) or because they fell outside the prioritized scope of theoretically motivated pairs (spatial–color, within-color, and geometric–geometric interactions).

Primary \ Secondary	Position	Length	Tilt	Area	Luminance	Saturation	Curvature
Length	✓		✓		✓	✓	✓
Tilt		✓		✓	✓	✓	✓
Area	✓		✓		✓	✓	
Luminance		✓	✓	✓		✓	✓
Saturation		✓	✓	✓	✓		✓
Curvature		✓	✓		✓	✓	

6 Separability

6.1 Task Design

The separability task assesses whether changes in one channel (the secondary) interfere with judgments about another (the primary).

The same 105 participants who completed the Accuracy experiment also completed the Separability experiment. For each tested primary–secondary channel pairing (Table 2), each participant performed a *magnitude-estimation judgment* (the same task as in Section 4) on the primary channel under two conditions: (i) the secondary fixed at its default (the Section 4 trials), and (ii) the secondary sampled uniformly across its full range (color channels varied in HSL). This added one response per pairing per participant, enabling within-subjects comparison of fixed versus varying secondary-channel effects. Rows in Table 2 denote the primary channel; columns denote the secondary channel. Position was tested only as a secondary channel; as a primary, it would largely replicate the length conditions given their shared spatial extent basis (Section 4).

6.2 Analysis

To quantify separability, we computed per-trial log-error using (Equation 1) with α=1, where P and S denote the participant's raw response and the true value, both normalized by the channel maximum $S_{\max}$.

Separability was assessed by the mean log-error difference between the varying and fixed secondary conditions: larger differences indicate stronger interference. We also tested the statistical significance of the difference relative to the baseline using paired t-tests with Benjamini–Hochberg FDR correction applied within each primary channel (FDR level 0.05), with the participant as the unit of analysis.

6.3 Results

Table 3 presents the tested separability matrix. Each cell reports the mean log-error (Equation 1, α=1, normalized to $[0,1]$ and thus negative) for the row (primary) channel judged while the column (secondary) channel varies. In each row, the gray cell with the underlined value is that row's *baseline*, sitting in its own-channel column (e.g., Length's baseline is in the Length column) and equal to its Section 4 accuracy; the remaining colored cells report performance while the column (secondary) channel varies. Dashed cells (–) are excluded pairs (Table 2). Cell color encodes the deviation from that baseline, $\Delta\text{error} = \text{cell} - \text{baseline}$ (blue: improvement, red: degradation, intensity $\propto |\Delta\text{error}|$ within each direction). Asterisks denote raw paired t-test significance, and bold cells survive BH–FDR correction.

The most pronounced effect is the **tilt–area** interaction. When area varies as a secondary channel, tilt judgment accuracy degrades from its baseline of -2.240 to -1.342 ($p < 0.001$, Hedges' $g = 1.285$), the largest effect across all tested pairings, bringing tilt performance near the chance level (≈ -1.33, the expected log-error under uniform

Table 3: Channel separability log-error matrix (mean). In each row, the gray cell with the underlined value marks that row's own-channel baseline (its Section 4 accuracy); the other cells report performance while the column (secondary) channel varies. Blue/red shading shows improvement/degradation vs. baseline (intensity ∝ magnitude within each direction). Asterisks mark raw paired t-test significance (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$); **bold** cells survive Benjamini–Hochberg FDR correction.

Primary Channel	Secondary Channel						
	Position	Length	Tilt	Area	Luminance	Saturation	Curvature
Length	**−2.253****	−2.086	−2.115	–	−2.001	−2.101	−2.090
Tilt	–	−2.110	−2.240	**−1.342*****	−2.111	−2.134	**−1.843*****
Area	−1.702	–	**−1.524***	−1.685	−1.600	−1.651	–
Luminance	–	**−1.765***	**−1.766***	**−1.843****	−1.590	−1.622	−1.626
Saturation	–	−1.766	−1.827	−1.722	**−1.545****	−1.843	−1.710
Curvature	–	−1.970	−1.913	–	−2.040	−1.927	−2.019

random responding). The reversed pairing (area judged as tilt varies) produces a smaller yet significant effect (−1.685 to −1.524; $p = 0.010$, Hedges' $g = 0.276$). Tilt also degrades substantially when paired with curvature (−2.240 to −1.843; $p < 0.001$, $g = 0.579$).

Among color channels, **luminance** variation disrupts **saturation** judgments (−1.843 to −1.545; $p < 0.01$, Hedges' $g = 0.487$), whereas saturation variation has minimal impact on luminance. Luminance judgments instead improve in several pairings when secondary channels vary (Table 3). Saturation degrades across most tested pairings.

Length judgments remain close to baseline across all tested secondary channels. Position variation improved length performance (−2.086 to −2.253; $p < 0.01$, $g = 0.299$).

6.4 Discussion

Three-tier interference hierarchy. Effect magnitudes reveal three interference tiers across the tested pairings.

- *Minimal observed interference.* Length with color channels shows the smallest deviations from baseline (within 0.085 log units), reflecting low interference rather than proven independence.
- *Moderate observed interference.* Curvature occupies an intermediate position, with log-errors ranging from −1.913 to −2.040 across tested partners (baseline: −2.019), indicating curvature is moderately robust to secondary-channel variation.
- *Strong interference.* Beyond the tilt–area case, the tilt–curvature and saturation–luminance pairings also produce significant performance degradation, with the latter exhibiting pronounced asymmetry. Position–length is a notable exception, where the varying position mark improves rather than degrades length judgments, indicating that spatial co-variation adds reference information.

Quantifying channel separability. While Garner's framework [24] established a clear boundary between separable and integral dimensions, our results reveal a continuum from mild to severe interference. The tilt–area pair exemplifies one extreme, where area variation fundamentally disrupts tilt processing. The observed asymmetry (e.g., luminance disrupts saturation but not vice versa) further challenges bidirectional separability assumptions, consistent with hierarchical visual processing in which earlier-processed luminance information interferes with higher-level color judgments [38].

We also observed facilitative effects. Luminance judgments improved with area variation, which likely reflects stimulus-level signal enhancement (larger area increases luminance extent) rather than a cross-channel mechanism; separating these accounts requires independent control of stimulus extent and encoded area value. The position–length facilitation, by contrast, allows a cleaner interpretation: the varying position mark provides an additional spatial reference, paralleling our accuracy finding that fixed reference frames improve length judgments (Section 4).

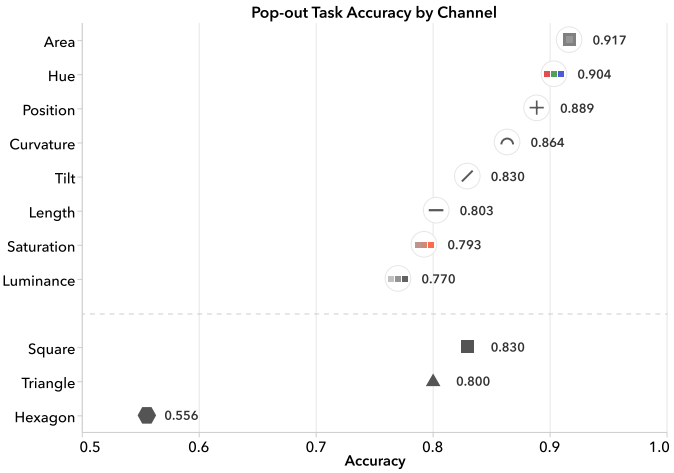

Fig. 5: Pop-out detection accuracy across visual channels under 100 ms exposure. Area and hue achieve the highest pop-out rates, while length shows only moderate pop-out effectiveness. Among shape channels the hexagon, the form most similar to the circular distractors, falls to near-chance detection, whereas more distinct forms remain salient.

7 Pop-out

7.1 Task Design

The pop-out task measures preattentive detection under brief exposure. In each trial, participants viewed an array of 20 elements (e.g., lines) on the screen, with or without a single outlier that differed in exactly one channel. Beyond the seven core channels, we additionally tested hue and shape (square, triangle, hexagon), which are relevant for pop-out but were excluded from the other three experiments because they lack ordinal magnitude scales.

Before each trial, the target channel was cued: participants were told which channel to monitor during the countdown, so search was guided rather than open-ended. Distractors were circles in the shape and hue conditions and the channel's default mark otherwise; array layout, distractor/outlier construction, and the resulting upper-bound interpretation appear in Supplementary S4. Following a 3-second countdown, the stimulus array was displayed for 100 ms, after which participants indicated whether an outlier was present and, if so, confirmed its location. We measured detection accuracy but not response time. The 100 ms exposure is consistent with established preattentive processing paradigms [9, 54]. Outlier values were standardized across the magnitude channels: distractors at 25% and outliers at 75% of the channel range (or vice versa), yielding a uniform separation of 50% of the channel range that exceeds the measured JND for every channel we tested in the discriminability task (Section 5). Outlier presence was randomized at 50%. We collected data from 105 participants with three trials per channel. Following singleton-search designs, our paradigm makes two deliberate simplifications: participants knew the target channel in

advance, and all distractors were identical [11, 41].

7.2 Results

Figure 5 presents pop-out detection accuracy across all tested channels, revealing a clear hierarchy with substantial variation both within and between channel categories. Among non-shape channels, **area** achieved the highest detection accuracy. **Hue** follows closely, while **position** also maintains strong performance. Geometric channels show moderate performance, with **curvature** and **tilt** achieving respectable but not exceptional detection rates.

Color channels exhibit striking variability in pop-out effectiveness. While **hue** achieves near-optimal performance, **saturation** and **luminance** show substantially lower detection rates. This disparity aligns with color vision research indicating that hue processing engages more fundamental visual pathways than saturation or brightness variations [38]. The 0.134 detection gap between hue and luminance represents one of the largest within-category differences observed, suggesting that not all color dimensions are equivalent for pop-out under brief exposure.

Length achieves only moderate pop-out performance despite its exceptional accuracy in quantitative estimation tasks. Crucially, this mismatch does not depend on area alone: even excluding area, length ranks fifth in detection performance. Area significantly outperforms length in pop-out detection (paired t-test: $p = 0.006$, Hedges' $g = 0.604$); part of this advantage may stem from stimulus extent or luminance cues, which we examine below.

Shape channels revealed a sharp drop from **square** and **triangle**, which maintained salience comparable to geometric channels, to **hexagon**, which fell to near-chance levels (paired t-test: $p < 0.001$, Hedges' $g = 1.029$). This pattern is consistent with the hexagon's close resemblance to the circular distractors; we return to the similarity-versus-complexity question below. Both contrasts remain significant under Bonferroni correction ($p = 0.012$ and $p < 0.002$) and permutation tests (10,000 iterations).

7.3 Discussion

Area's detection advantage contradicts accuracy-based expectations. Area achieved the highest pop-out detection accuracy despite requiring a power-law correction for quantitative reading. Two factors may contribute. Area has the finest discriminability (lowest JND) over most of its range (Figure 4), so the uniform half-range separation is a larger suprathreshold ratio than for other channels. An area outlier also occupies a larger physical footprint, which enhances attentional capture [43] and adds a luminance-contrast cue absent where outlier and distractor extent match. Our data cannot separate whether area's advantage reflects rapid size processing, low-level luminance contrast, or both. For visualization design, this finding applies mainly to displays with uniform distractors, where area is a strong channel for anomaly highlighting. In practice, area often already encodes a data variable (e.g., bubble size), limiting its availability for highlighting.

Dissociation between accuracy and pop-out. Length, by contrast, shows only moderate pop-out despite top-tier accuracy (fifth even when area is excluded), indicating that detection and estimation precision are governed by distinct mechanisms not captured by accuracy-based rankings alone [13, 30].

Color hierarchy and shape similarity. Hue's superior pop-out, plausibly mediated by opponent-color mechanisms [38], suggests prioritizing hue over luminance or saturation for highlighting when pop-out is the goal. The three tested shapes jointly vary geometric complexity and similarity to the circular distractors, so our design cannot separate the two accounts; the hexagon, the polygon most similar to the distractors, was the hardest to detect, suggesting that a more distinctive shape might pop out even if geometrically complex [25]. These patterns support Feature Integration Theory while revealing that primitive-stimulus contexts can reduce the effectiveness of features relying on contextual contrast (e.g., tilt) and amplify those that remain effective under brief, guided viewing (e.g., area).

8 General Discussion

8.1 A Scenario-driven Perspective on Channel Effectiveness

Prior studies on visualization effectiveness typically conclude with design implications for specific chart types and tasks, but existing frameworks do not sufficiently account for the **user's scenario**. Here, the scenario is distinct from visualization users or tasks such as those taxonomized by Amar et al. [1] or Brehmer et al. [8], in that the same abstract task can arise under different circumstances. Rather than inheriting encoding choices prescribed by visualization type or task category, we encourage designers to break from this convention and proactively identify the channel that best fits their scenario. Based on our findings, we organize the design implications into four recurring dimensions of scenario (Task Stakes, Data Granularity, Background Interference, and Response Time), each mapped to a corresponding task (accuracy, discriminability, separability, and pop-out).

Task Stakes corresponds to the *accuracy* task, reflecting how precisely a value must be extracted from an encoded channel. A battery indicator in mobile phones tolerates coarse readout, whereas a scientific figure or finance dashboard may require much more precise extraction. In low-stakes contexts, a designer may therefore favor lower-accuracy channels for aesthetics or space efficiency, reserving high-accuracy channels, like position along a common scale, for contexts where precise readout is critical.

Data Granularity maps to *discriminability*, capturing how finely the data must be expressed and interpreted by the viewer. When data is left-skewed with values concentrated at the higher end of the range, saturation may be a poor choice, as it offers lower discriminability in that region. One mitigation strategy is selective sampling: for the Tilt channel spanning 0°–180°, sampling near the cardinal orientations (0°, 90°, 180°) exploits the angles where discriminability is sharpest. Conversely, right-skewed data (such as a Pareto distribution) maps naturally onto the lower range, where saturation is more discriminable.

Background Interference arises when a visual channel must remain perceptible against competing visual elements, which relates to *separability* and becomes especially relevant when charts are embellished to enhance engagement. For instance, encoding data values through color hue becomes problematic when the background itself is richly colored or textured, a common occurrence in sports infographics or editorial data journalism, whereas position or length channels tend to remain separable under heavier visual embellishment. Because our separability experiment measures pairwise channel interference, not background decoration directly, we treat this dimension as a design extrapolation from that pairwise evidence rather than a tested claim; the principle is that concurrent visual variation degrades target-channel judgment.

Response Time governs how quickly a value must be retrieved or acknowledged, which relates to *pop-out*. Consider a digital car dashboard where vehicle speed is rendered as a large numeral at the center. Though the speed value itself is not encoded through a conventional perceptual channel, its relative importance is communicated through size, causing it to pop out from the surrounding display before the driver consciously searches for it. A similar principle applies to alert systems and real-time monitoring dashboards, where color, particularly red, can serve as a preattentive signal that draws the viewer's attention to a critical threshold even before any deliberate search begins.

Because rankings shift across tasks and a scenario is shaped by a broad space of design, data, user, and environmental factors, our scenario-driven perspective provides a practical, non-prescriptive guide that enables designers to make principled channel selections grounded in the conditions under which a visualization is actually consumed.

8.2 Convergence and Divergence with Chart-Level Evidence

We compare our results with published chart-level studies to distinguish convergences from context-dependent divergences.

8.2.1 Convergences

Several key findings align with chart-level evidence. The tilt–area interference parallels studies showing that area dominates angle in

pie charts [35, 46], and the luminance–saturation asymmetry mirrors directional separability in scatterplots [45, 47]. The reference-frame advantage also helps explain the accuracy gap between aligned and stacked bar charts [30, 52], while area's power-law correction replicates value underestimation in treemaps [34]. Finally, the cross-task ranking shifts align with prior evidence of task-dependent channel effectiveness [40, 44].

8.2.2 Divergences

Angle perception in chart contexts. Tilt falls within the pre-specified equivalence margin (± 0.2 log-error) of single position in our primitive stimuli, yet ranks among the worst in pie chart studies [13, 30]. This divergence arises from *channel confusion*: pie charts co-vary angle, area, and arc length, and viewers rely on the more salient area cue [35, 46]. Our result captures isolated orientation judgment, distinct from the angular extent of pie wedges, while chart-level studies capture the *effective* channel actually used in multi-cue displays. Therefore, angle encodings should not be dismissed as inherently imprecise, but designers must ensure that competing cues do not override the intended channel.

Pop-out under distractor heterogeneity. Area achieves the highest pop-out detection in our paradigm (0.917), where all non-target elements are identical. However, preattentive detection degrades with heterogeneous distractors [27]. Our measurements, therefore, represent an upper bound, strongest when non-target marks are uniform (small-multiple designs) and weakest in dense, heterogeneous displays.

Scaffolding amplifies spatial channels. Gridlines and axes disproportionately benefit spatial channels [13], so the accuracy gap between spatial and non-spatial channels is *wider* in scaffolded contexts than in our primitive stimuli. Conversely, in scaffolding-free contexts (e.g., augmented reality [22], embedded visualizations [58], data physicalizations [19]), our measurements may better predict channel performance than chart-level rankings.

These comparisons suggest that primitive stimulus and chart-level paradigms are *complementary*: convergences (tilt–area interference, baseline effects, area's power law) point to shared perceptual structure, while divergences occur precisely where theory predicts, namely competing cues (tilt in pie charts), heterogeneous distractors (pop-out), and spatial scaffolding (gridlines for position). Recognizing these three modulating factors enables designers to predict *when* our findings transfer directly (e.g., scaffolding-free AR or embedded visualizations) and *when* context-specific adjustments are necessary.

8.3 Limitations

Our study relied on participants recruited via Mechanical Turk, introducing uncontrolled variation in display hardware, ambient lighting, and screen calibration. Attention checks mitigated this, though residual noise in color-based channels is possible.

Also, our experiments prioritized breadth over depth, using sparse per-participant trial counts (two responses for accuracy, one per pairing for separability, and three for pop-out). Evaluating channels across four perceptual tasks and many channel combinations limited how finely we could probe each domain, particularly for separability, display density, and parametric variation in channel magnitudes. This breadth-first design yields broad multi-task profiles through large group-level aggregation ($N = 105$), but lacks individual-level reliability estimation. Higher trial counts per participant would be needed in follow-up studies to characterize individual differences [15]. Sparse per-participant data inflate the standard error of paired differences, so our tests are conservative and the reported effects survive nonetheless. We randomized task-block order across participants to control for learning and fatigue from the repeated blocks, though residual effects remain possible.

Several findings lack a clear theoretical explanation. Most notably, luminance showed increased accuracy in all tested separability conditions when paired with an interfering channel. These effects were small to moderate and their mechanism remains unconfirmed; characterizing them with targeted replication is future work.

The shared canvas frame was a deliberate constant across stimuli, and its influence is most visible in the strong right-boundary anchoring of length and area in discriminability. These boundary effects may not transfer to frameless displays.

Finally, we note limitations of each perceptual task. In terms of accuracy, judgment ranges were defined by channel-specific geometry or by the pre-block reference rather than uniformly by the canvas, and these references provide unequal perceptual support across channels. For length and area, the canvas boundary can serve as a direct comparison reference, while position is read against its own segment and the color channels rely on the pre-block reference. The measured accuracy advantage of spatial over color channels partly reflects reference-frame availability rather than intrinsic perceptual superiority alone.

For separability, our design was more sensitive to moderate effects than to small ones, with roughly $N \approx 105$ paired observations per comparison. Accordingly, channel pairs showing no significant difference should not be interpreted as evidence of separability: small effects may remain undetected after FDR correction, and the "minimal observed interference" tier reflects the absence of detected interference rather than confirmed independence.

Position was excluded from discriminability and from the separability primary channel under the assumption that side-by-side comparison neutralizes the reference-frame advantage distinguishing position from length in absolute estimation (Section 4, $p = 0.009$), consistent with distinctions between absolute identification and relative judgment tasks [50]. This assumption is plausible where both reduce to spatial-extent comparison under simultaneous viewing, but was not tested; future work should verify it by including position in discriminability.

Saturation was tested with a single base hue (red, hue $= 0°$); generalization to other hues requires further investigation, as saturation discrimination may vary across the hue circle. We also tested all channels within normalized ranges (0–100 for metric channels and 0°–180° for tilt and curvature); perceptual behavior may differ when values are sampled from narrower subranges or different portions.

9 Conclusion

Our paper presents a systematic evaluation of visual channels across four critical perceptual tasks using primitive visual stimuli that remove chart-specific scaffolding while retaining a shared display frame.

Our findings demonstrate that channel effectiveness is fundamentally multi-dimensional; channels that excel in one task may falter in another, and tested pairwise interactions can override individual channel strengths. Rather than proposing another unified ranking, our work provides multi-task profiles, separability constraints, and cross-paradigm comparisons, and organizes these findings through a scenario-driven perspective around four dimensions of visualization usage (task stakes, data granularity, background interference, and response time) for context-sensitive channel selection.

Practically, interference-based constraints such as avoiding tilt–area combinations are among our most robust findings and serve as concrete design warnings within the tested setting. Task-specific channel selection that uses high-salience channels for preattentive detection and high-accuracy channels for precise reading offers a principled alternative to one-size-fits-all guidelines. These principles can also complement automated recommendation systems [39, 60] through multi-dimensional channel profiles rather than single-metric rankings.

We also suggest several directions for expanding our approach. Systematic validation with richer stimuli would clarify the boundary conditions of our findings, particularly for the divergences identified in Section 8. Controlled follow-up on facilitative effects (e.g., position–length improvement) may reveal deliberate channel pairings that enhance rather than merely preserve perception. Finally, extending the separability framework to three-way interactions and individual differences could enable personalized visualization recommendations and reader-adaptive interpretation support [12]. As visualization moves beyond traditional charts to more complex forms, such as augmented reality, embedded graphics, and data physicalizations, these fundamental channel-level insights become increasingly critical for effective visualization design.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2023R1A2C2005209), the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) [No. RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)], the AI Seoul Tech Research Support Program of the Seoul Future Foundation, and the SNU-Global Excellence Research Center establishment project. The ICT at Seoul National University provided research facilities for this study.

References

[1] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS 2005)*, pp. 111–117. IEEE Computer Society, Los Alamitos, 2005. doi: 10.1109/INFVIS.2005.1532136 8

[2] L. Bartram, B. Cheung, and M. Stone. The effect of colour and transparency on the perception of overlaid grids. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1942–1948, 2011. doi: 10.1109/TVCG.2011.242 2

[3] C. X. Bearfield, C. Stokes, A. Lovett, and S. Franconeri. What does the chart say? Grouping cues guide viewer comparisons and conclusions in bar charts. *IEEE Transactions on Visualization and Computer Graphics*, 30(8):5097–5110, 2024. doi: 10.1109/TVCG.2023.3289292 2

[4] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x 4

[5] J. Bertin. *Semiology of graphics*. University of Wisconsin Press, Madison, 1983. 2

[6] A. Bezerianos and P. Isenberg. Perception of visual variables on tiled wall-sized displays for information visualization applications. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2516–2525, 2012. doi: 10.1109/TVCG.2012.251 3

[7] H. R. Blackwell. Contrast thresholds of the human eye. *Journal of the Optical Society of America*, 36(11):624–643, 1946. doi: 10.1364/JOSA.36.000624 6

[8] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013. doi: 10.1109/TVCG.2013.124 8

[9] D. Bridges, A. Pitiot, M. R. MacAskill, and J. W. Peirce. The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, 8:e9414, 2020. doi: 10.7717/peerj.9414 7

[10] F. W. Campbell, J. J. Kulikowski, and J. Levinson. The effect of orientation on the visual resolution of gratings. *The Journal of Physiology*, 187(2):427–436, 1966. doi: 10.1113/jphysiol.1966.sp008100 2, 5, 6

[11] Z. Chen and K. R. Cave. Singleton search is guided by knowledge of the target, but maybe it shouldn't be. *Vision Research*, 115:92–103, 2015. doi: 10.1016/j.visres.2015.08.012 8

[12] K. Choe, C. Lee, S. Lee, J. Song, A. Cho, N. W. Kim et al. Enhancing data literacy on-demand: LLMs as guides for novices in chart interpretation. *IEEE Transactions on Visualization and Computer Graphics*, 31(9):4712–4727, 2025. doi: 10.1109/TVCG.2024.3413195 9

[13] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984. doi: 10.1080/01621459.1984.10478080 1, 2, 4, 8, 9

[14] Z. Cutler, J. Wilburn, H. Shrestha, Y. Ding, B. Bollen, K. A. Nadib et al. reVISit 2: A full experiment life cycle user study framework. *IEEE Transactions on Visualization and Computer Graphics*, 32(1):13–23, 2026. doi: 10.1109/TVCG.2025.3633896 3

[15] R. Davis, X. Pu, Y. Ding, B. D. Hall, K. Bonilla, M. Feng et al. The risks of ranking: Revisiting graphical perception to model individual differences in visualization performance. *IEEE Transactions on Visualization and Computer Graphics*, 30(3):1756–1771, 2024. doi: 10.1109/TVCG.2022.3226463 9

[16] Ç. Demiralp, M. S. Bernstein, and J. Heer. Learning perceptual kernels for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1933–1942, 2014. doi: 10.1109/TVCG.2014.2346978 2

[17] M. Dick and S. Hochstein. Visual orientation estimation. *Perception & Psychophysics*, 46(3):227–234, 1989. doi: 10.3758/BF03208083 2

[18] Y. Ding, J. Wilburn, H. Shrestha, A. Ndlovu, K. Gadhave, C. Nobre et al. reVISit: Supporting scalable evaluation of interactive visualizations. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pp. 31–35. IEEE Computer Society, Los Alamitos, 2023. doi: 10.1109/VIS54172.2023.00015 3

[19] P. Dragicevic, Y. Jansen, and A. Vande Moere. Data physicalization. In J. Vanderdonckt, P. Palanque, and M. Winckler, eds., *Handbook of Human Computer Interaction*, pp. 1–51. Springer International Publishing, Cham, 2021. doi: 10.1007/978-3-319-27648-9_94-1 9

[20] G. T. Fechner. *Elements of Psychophysics*, vol. 1. Holt, Rinehart and Winston, New York, 1966. Original work published 1860. 2

[21] J. J. Flannery. The relative effectiveness of some common graduated point symbols in the presentation of quantitative data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 8(2):96–109, 1971. doi: 10.3138/J647-1776-745H-3667 4

[22] A. Fonnet and Y. Prié. Survey of immersive analytics. *IEEE Transactions on Visualization and Computer Graphics*, 27(3):2101–2122, 2021. doi: 10.1109/TVCG.2019.2929033 9

[23] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman. The science of visual data communication: What works. *Psychological Science in the Public Interest*, 22(3):110–161, 2021. doi: 10.1177/15291006211051956 2

[24] W. R. Garner. *The Processing of Information and Structure*. The Experimental Psychology Series. Lawrence Erlbaum Associates, Potomac, 1974. doi: 10.4324/9781315802862 2, 7

[25] M. Grüner, F. Goller, and U. Ansorge. Simple shapes guide visual attention based on their global outline or global orientation contingent on search goals. *Journal of Experimental Psychology: Human Perception and Performance*, 47(11):1493–1515, 2021. doi: 10.1037/xhp0000955 8

[26] D. Haehn, J. Tompkin, and H. Pfister. Evaluating 'graphical perception' with CNNs. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):641–650, 2019. doi: 10.1109/TVCG.2018.2865138 2

[27] S. Haroz and D. Whitney. How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2402–2410, 2012. doi: 10.1109/TVCG.2012.233 9

[28] L. Harrison, F. Yang, S. Franconeri, and R. Chang. Ranking visualizations of correlation using Weber's law. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1943–1952, 2014. doi: 10.1109/TVCG.2014.2346979 4

[29] C. Healey and J. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, 2012. doi: 10.1109/TVCG.2011.127 2

[30] J. Heer and M. Bostock. Crowdsourcing graphical perception: using Mechanical Turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pp. 203–212. Association for Computing Machinery, New York, 2010. doi: 10.1145/1753326.1753357 1, 2, 3, 4, 8, 9

[31] M. Kay and J. Heer. Beyond Weber's law: A second look at ranking visualizations of correlation. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):469–478, 2016. doi: 10.1109/TVCG.2015.2467671 5

[32] D. Kiesel, P. Riehmann, and B. Froehlich. Rose charts: Area or length encoding for fill level of circle sectors? In *EuroVis 2025 - Short Papers*. The Eurographics Association, Goslar, 2025. doi: 10.2312/evs.20251075 1

[33] Y. Kim and J. Heer. Assessing effects of task and data distribution on the effectiveness of visual encodings. *Computer Graphics Forum*, 37(3):157–167, 2018. doi: 10.1111/cgf.13409 2

[34] N. Kong, J. Heer, and M. Agrawala. Perceptual guidelines for creating rectangular treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):990–998, 2010. doi: 10.1109/TVCG.2010.186 9

[35] R. Kosara. Evidence for area as the primary visual cue in pie charts. In *2019 IEEE Visualization Conference (VIS)*, pp. 101–105. IEEE Computer Society, Los Alamitos, 2019. doi: 10.1109/VISUAL.2019.8933547 9

[36] S. Lee, M. Chang, S. Park, and J. Seo. Assessing graphical perception of image embedding models using channel effectiveness. In *2024 IEEE Visualization and Visual Analytics (VIS)*, pp. 226–230. IEEE Computer Society, Los Alamitos, 2024. doi: 10.1109/VIS55277.2024.00053 2, 3

[37] S. Lee, J. Kim, S. Park, S. Lee, J. Song, B. Kim et al. Disentangling visual and factual correctness in LVLMs' visualization literacy. *arXiv preprint arXiv:2606.03142*, 2026. doi: 10.48550/arXiv.2606.03142 2

[38] M. Livingstone and D. Hubel. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, 240(4853):740–749, 1988. doi: 10.1126/science.3283936 6, 7, 8

[39] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141, 1986. doi: 10.1145/22949.22950 1, 2, 9

[40] C. M. McColeman, F. Yang, T. F. Brady, and S. Franconeri. Rethinking the ranks of visual channels. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):707–717, 2022. doi: 10.1109/TVCG.2021.3114684 2, 9

[41] H. J. Müller, D. Heller, and J. Ziegler. Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, 57(1):1–17, 1995. doi: 10.3758/BF03211845 8

[42] T. Munzner. *Visualization analysis and design*. CRC Press, Boca Raton, 2014. doi: 10.1201/b17511 2, 3

[43] M. J. Proulx. Size matters: large objects capture attention in visual search. *PLOS ONE*, 5(12):e15293, 2010. doi: 10.1371/journal.pone.0015293 8

[44] B. Saket, A. Endert, and Ç. Demiralp. Task-based effectiveness of basic visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(7):2505–2512, 2019. doi: 10.1109/TVCG.2018.2829750 2, 9

[45] K. B. Schloss, C. C. Gramazio, A. T. Silverman, M. L. Parker, and A. S. Wang. Mapping color to meaning in colormap data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):810–819, 2019. doi: 10.1109/TVCG.2018.2865147 9

[46] D. Skau and R. Kosara. Arcs, angles, or areas: Individual data encodings in pie and donut charts. *Computer Graphics Forum*, 35(3):121–130, 2016. doi: 10.1111/cgf.12888 1, 9

[47] S. Smart and D. A. Szafir. Measuring the separability of shape, size, and color in scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 1–14. Association for Computing Machinery, New York, 2019. doi: 10.1145/3290605.3300899 2, 9

[48] I. Spence and S. Lewandowsky. Displaying proportions and percentages. *Applied Cognitive Psychology*, 5(1):61–77, 1991. doi: 10.1002/acp.2350050106 4

[49] S. S. Stevens. On the psychophysical law. *Psychological Review*, 64(3):153–181, 1957. doi: 10.1037/h0046162 2, 4

[50] N. Stewart, G. D. Brown, and N. Chater. Absolute identification by relative judgment. *Psychological Review*, 112(4):881–911, 2005. doi: 10.1037/0033-295X.112.4.881 9

[51] D. A. Szafir. Modeling color difference for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):392–401, 2018. doi: 10.1109/TVCG.2017.2744359 2

[52] J. Talbot, V. Setlur, and A. Anand. Four experiments on the perception of bar charts. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2152–2160, 2014. doi: 10.1109/TVCG.2014.2346320 9

[53] R. Teghtsoonian. On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, 78(1):71–80, 1971. doi: 10.1037/h0030300 4

[54] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980. doi: 10.1016/0010-0285(80)90005-5 2, 7

[55] R. Veras and C. Collins. Discriminability tests for visualization effectiveness and scalability. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):749–758, 2020. doi: 10.1109/TVCG.2019.2934432 2

[56] M. Waldner, A. Diehl, D. Gračanin, R. Splechtna, C. Delrieux, and K. Matković. A comparison of radial and linear charts for visualizing daily patterns. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1033–1042, 2020. doi: 10.1109/TVCG.2019.2934784 1

[57] G. Westheimer. Meridional anisotropy in visual processing: implications for the neural site of the oblique effect. *Vision Research*, 43(22):2281–2289, 2003. doi: 10.1016/S0042-6989(03)00360-2 2

[58] W. Willett, Y. Jansen, and P. Dragicevic. Embedded data representations. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):461–470, 2017. doi: 10.1109/TVCG.2016.2598608 9

[59] J. M. Wolfe and T. S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6):495–501, 2004. doi: 10.1038/nrn1411 2

[60] K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, and J. Heer. Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):649–658, 2016. doi: 10.1109/TVCG.2015.2467191 9

SUPPLEMENTARY MATERIAL

S1 Discriminability: Model Comparison and Contour Details

To validate the Anchored Harmonic Weber model (Section 5), we compared four nested models of increasing complexity. Each successive model adds parameters that test a distinct perceptual hypothesis:

1. **Null** (constant JND):

$$|\Delta|(x) = \text{offset}_i$$

   Parameters: offset_i per contour level ($k = 5$, corresponding to the five "Same"-response thresholds at 10%–50%).

2. **Weber** (JND proportional to stimulus):

$$|\Delta|(x) = w_0 \cdot \frac{x}{x_{max}} + \text{offset}_i$$

   Adds one shared slope w_0 ($k = 6$). Tests whether discriminability scales with stimulus magnitude.

3. **Symmetric anchored** ($w_L = w_R = w$):

$$|\Delta|(x) = w_0 \cdot \frac{x}{x_{max}} + w \cdot \frac{x\,(x_{max} - x)}{x_{max}} + \text{offset}_i$$

   Adds one shared boundary weight w ($k = 7$). The parabolic term captures improved discriminability near both domain boundaries, but constrains the effect to be symmetric.

4. **Full anchored harmonic** (asymmetric $w_L \neq w_R$):

$$|\Delta|(x) = w_0 \cdot \frac{x}{x_{max}} + 1/\left(\frac{1}{w_L x} + \frac{1}{w_R(x_{max} - x)}\right) + \text{offset}_i$$

   Shared parameters w_0, w_L, w_R ($k = 8$). Allows each boundary to exert a different anchoring strength.

All models share the same contour-level offsets (offset_i, $i = 1, \ldots, 5$); the models differ only in their shared shape parameters. For each model, we minimized the sum of squared residuals (RSS) against the extracted discriminability contours and computed BIC $= n\ln(RSS/n) + k\ln n$, where n is the number of data points and k is the total parameter count.

Table S1 reports the full comparison. Across all six channels (with tilt split into two subranges, yielding seven fitted conditions), the full anchored harmonic model achieves the lowest BIC ($\Delta\text{BIC} \geq 52$ relative to the next-best model), confirming that asymmetric dual-boundary anchoring captures perceptual structure beyond what simpler models explain.

Contour extraction. The fitted contours in Figure 3 are iso-probability contours of each channel's response-probability grid (the proportion of "Same" responses over the reference-value × $|\Delta|$ space), to which the Anchored Harmonic Weber model is fit. Where the fitted model dips below zero near a domain boundary, this reflects extrapolation of the parametric fit rather than a negative threshold, as the JND itself is non-negative.

Tilt left–right asymmetry. We fit the half-ranges 0–90° and 90–180° separately because the cardinal orientations act as high-sensitivity anchors. Within each half the fitted right-end slope exceeds the left-end slope (0–90°: 1.030 vs. 0.120; 90–180°: 0.986 vs. 0.149; Table 1): discrimination sharpens steeply approaching 90° from below and 180°, yet only weakly approaching 90° from above, so the 90° anchor is asymmetric by approach direction. We report this as an empirical pattern: orientation was sampled uniformly and the design was not built to dissect the two sides of the 90° anchor, so denser sampling near the cardinal axes is needed to confirm it.

Table S1: Nested model comparison for discriminability. For each channel, four models of increasing complexity are compared. **Bold** indicates the best (lowest) BIC per channel. k: total parameters; n: data points; RMSE: root mean squared error.

Channel	Model	k	n	RSS	RMSE	R^2	BIC
Area	Null	5	537	1618.81	1.736	0.517	623.98
	Weber	6	537	579.01	1.038	0.827	78.16
	Symmetric	7	537	253.51	0.687	0.924	−359.08
	Full anchored	**8**	**537**	**227.33**	**0.651**	**0.932**	**−411.31**
Curvature	Null	5	1082	20991.73	4.405	0.159	3243.41
	Weber	6	1082	3430.45	1.781	0.863	1290.42
	Symmetric	7	1082	3241.88	1.731	0.870	1236.23
	Full anchored	**8**	**1082**	**1071.17**	**0.995**	**0.956**	**45.01**
Length	Null	5	575	3781.84	2.565	0.353	1114.84
	Weber	6	575	1798.18	1.768	0.692	693.72
	Symmetric	7	575	923.30	1.267	0.842	316.79
	Full anchored	**8**	**575**	**650.61**	**1.064**	**0.888**	**121.87**
Luminance	Null	5	571	1874.90	1.812	0.454	710.60
	Weber	6	571	1874.88	1.812	0.454	716.96
	Symmetric	7	571	930.92	1.277	0.729	323.53
	Full anchored	**8**	**571**	**229.21**	**0.634**	**0.933**	**−470.40**
Saturation	Null	5	580	6580.23	3.368	0.181	1440.52
	Weber	6	580	217.34	0.612	0.973	−531.14
	Symmetric	7	580	213.05	0.606	0.974	−536.34
	Full anchored	**8**	**580**	**104.47**	**0.424**	**0.986**	**−943.29**
Tilt (0°–90°)	Null	5	558	2731.33	2.212	0.366	917.83
	Weber	6	558	1560.66	1.672	0.638	611.85
	Symmetric	7	558	938.50	1.297	0.782	334.39
	Full anchored	**8**	**558**	**723.10**	**1.138**	**0.827**	**195.22**
Tilt (90°–180°)	Null	5	610	3565.77	2.418	0.602	1109.13
	Weber	6	610	3303.59	2.327	0.632	1068.96
	Symmetric	7	610	1391.62	1.510	0.845	548.00
	Full anchored	**8**	**610**	**1162.40**	**1.380**	**0.867**	**444.62**

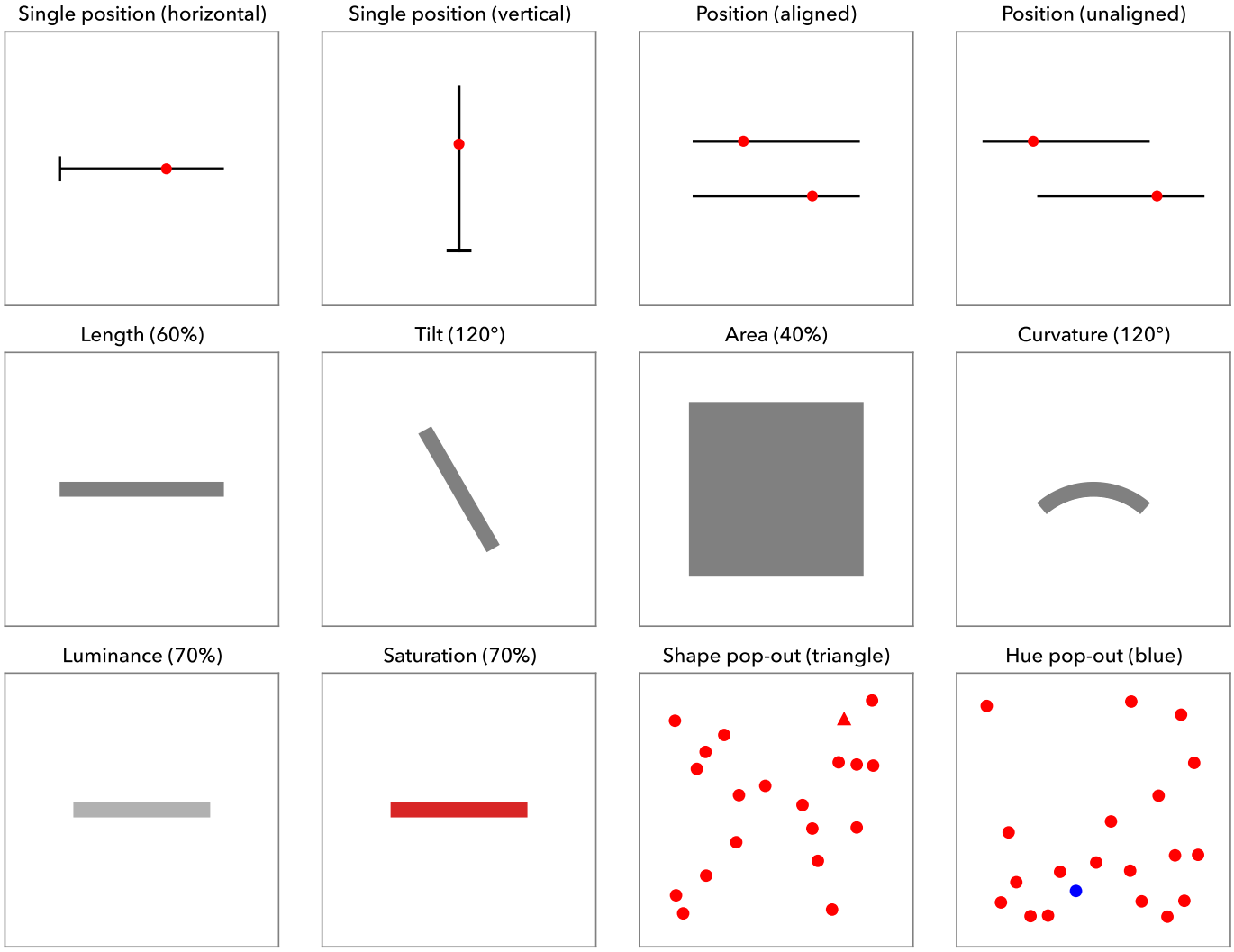

Fig. S1: Representative stimuli for each channel and condition. Top row: the three position variants (single, aligned, unaligned); single is shown in both horizontal and vertical orientations, and the red dot marks the point to be read. Middle row: length, tilt, area, and curvature. Bottom row: the achromatic luminance line and the red saturation line (both rendered as colored line marks, not filled squares), and the pop-out arrays for shape (20 elements, an outlier triangle among identical red circles) and hue (an outlier blue circle among red). Marks are enlarged slightly for print visibility; per-channel value ranges are in Section 3.1.

S2 Stimulus Examples

Figure S1 shows a representative stimulus for each channel and condition; per-channel value ranges and increments are given in Section 3.1, and all marks sit on the 500×500 px canvas.

S3 Analysis Details

S3.1 The $1/8$ Offset and Its Sensitivity

The additive offset $\frac{1}{8} = 0.125$ inside the logarithm in Equation 1 is adapted from the log-error measure of Cleveland and McGill and prevents the logarithm from diverging when a response exactly matches the ground truth ($|P - S^{\alpha}| = 0$). With it, a perfectly correct response maps to $\log_2(0.125) = -3$, the best-case floor on the mean-log-error scale, and chance-level responding under a uniform model corresponds to ≈ -1.33. To confirm that our conclusions are not an artifact of this choice, we recomputed the per-channel accuracy baselines and the channel ordering under alternative offsets (Table S2). The ordering is essentially invariant relative to the reference offset $1/8$ (Spearman $\rho = 1.00$ for $1/16$ and 0.97 for $1/4$ and $1/2$; only the adjacent single-position/tilt and aligned-position/curvature pairs swap at the two largest offsets). The headline single-position vs. length comparison remains significant for offsets $1/16$ ($p = 0.002$), $1/8$ ($p = 0.009$), and $1/4$ ($p = 0.039$), but not for the largest tested offset, $1/2$ ($p = 0.12$).

Table S2: Offset sensitivity: per-channel accuracy baseline mean log-error m under alternative offsets, in ranked order. The ordering is essentially unchanged (Spearman $\rho \geq 0.97$ vs. the $1/8$ reference).

Channel	$1/16$	$1/8$	$1/4$	$1/2$
Position (single)	-2.849	-2.240	-1.524	-0.718
Tilt	-2.805	-2.240	-1.545	-0.742
Length	-2.597	-2.086	-1.442	-0.680
Position (aligned)	-2.533	-2.029	-1.396	-0.646
Curvature	-2.499	-2.019	-1.400	-0.656
Saturation	-2.268	-1.843	-1.278	-0.578
Position (unaligned)	-2.221	-1.791	-1.230	-0.540
Area	-2.054	-1.685	-1.171	-0.513
Luminance	-1.942	-1.590	-1.097	-0.460

S3.2 Normality and Nonparametric Robustness

Shapiro–Wilk tests did not reject normality for 29 of the 36 Accuracy paired-difference distributions, and Wilcoxon signed-rank checks

agreed with the headline Accuracy comparisons. Separately, for the bounded, few-trial Pop-out proportions, where normality is least plausible, the two principal contrasts were verified with sign-flip permutation tests (10,000 iterations).

S3.3 Pooling of the Two Discriminability Groups

Discriminability used two groups (105 core at 10 trials per channel; 45 additional at 32), merged and pooled at the *trial* level: the response-probability grid averages the "Same" indicator over all trials in each (reference value, $|\Delta|$) cell, so every trial contributes equally and participants with 32 trials therefore contribute proportionally more observations to a cell than those with 10. The unequal allocation was an intentional design choice: the additional trials densify coverage of the reference-value $\times$ offset space for contour extraction. A Mann–Whitney comparison of the two groups' per-participant "Same" rates found no detectable difference on five of the six channels; saturation was the exception, so its contours rest more heavily on the pooling assumption. As a stronger check, we refit the Anchored Harmonic Weber model on the core sample alone (10 trials per channel) and compared it with the full-sample fit. The channel discriminability ordering is identical across the two (Spearman $\rho = 1.0$) and the per-channel parameters are consistent, which supports the robustness of the reported ordering, though not full equivalence between the two groups.

S4 Pop-out: Array Layout and Cueing Scope

Figure S1 (bottom row) shows the 20-element pop-out arrays for the shape and hue conditions with the outlier present; distractors are identical within an array (in these two conditions, red circles), so the shape condition pits an outlier shape against circular distractors and the hue condition a single blue circle against red ones. The positions of the 20 elements were scattered pseudo-randomly within the display. Before each trial a text prompt named the single channel to monitor, shown during the 3-second countdown, so participants knew which feature dimension to attend before the 100 ms array appeared. Combined with the uniform-distractor design, this advance cueing makes the task a guided singleton search: it favors detection and represents an upper bound on preattentive salience rather than pure bottom-up pop-out, where the relevant dimension is unknown and distractors are heterogeneous. Our detection rates should therefore be read as channel-level salience under guided, uniform-distractor conditions; performance in unguided, heterogeneous displays may be lower (Section 7, Section 8).