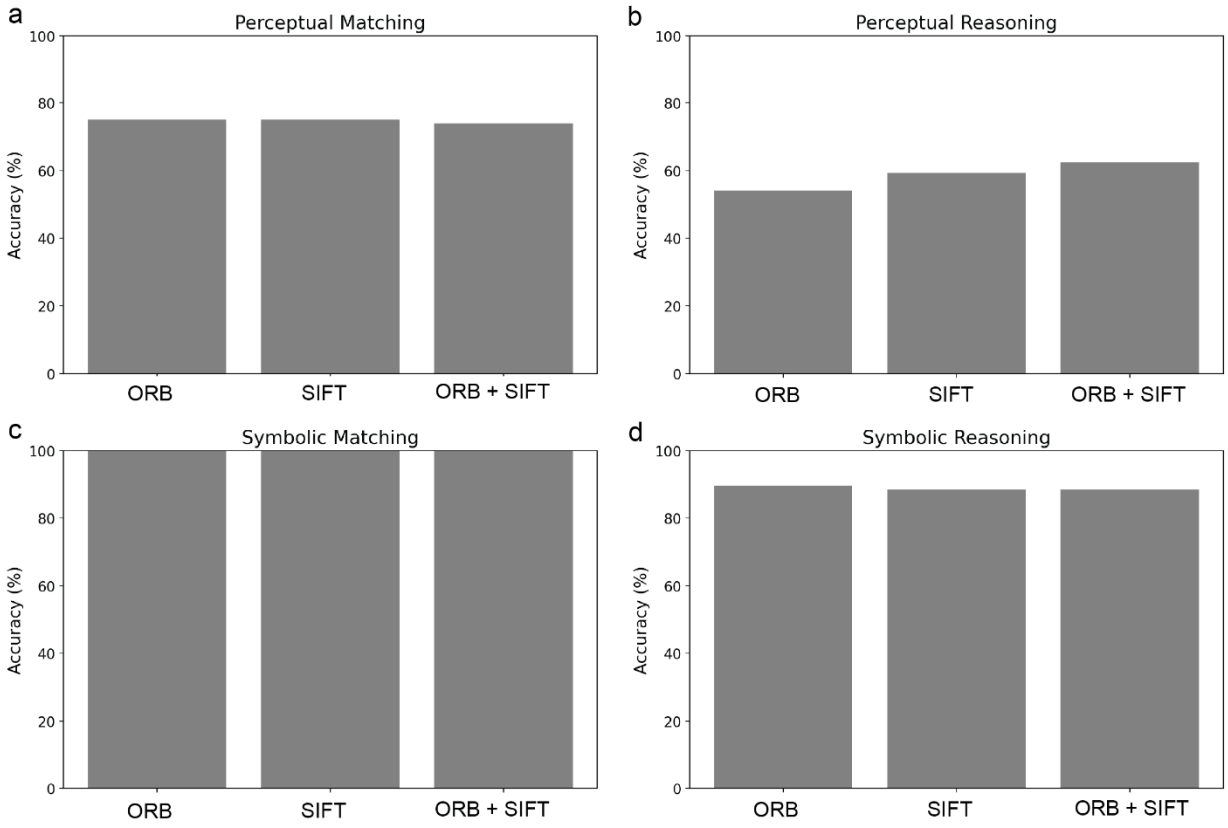
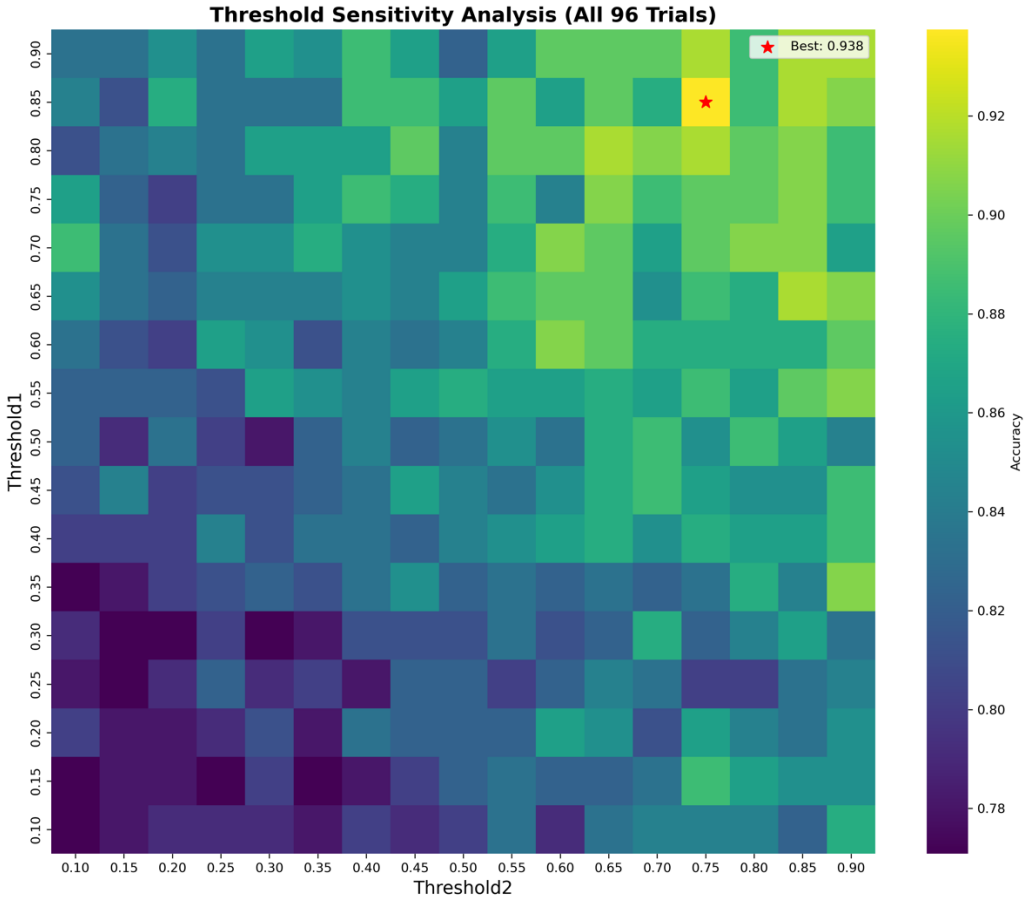
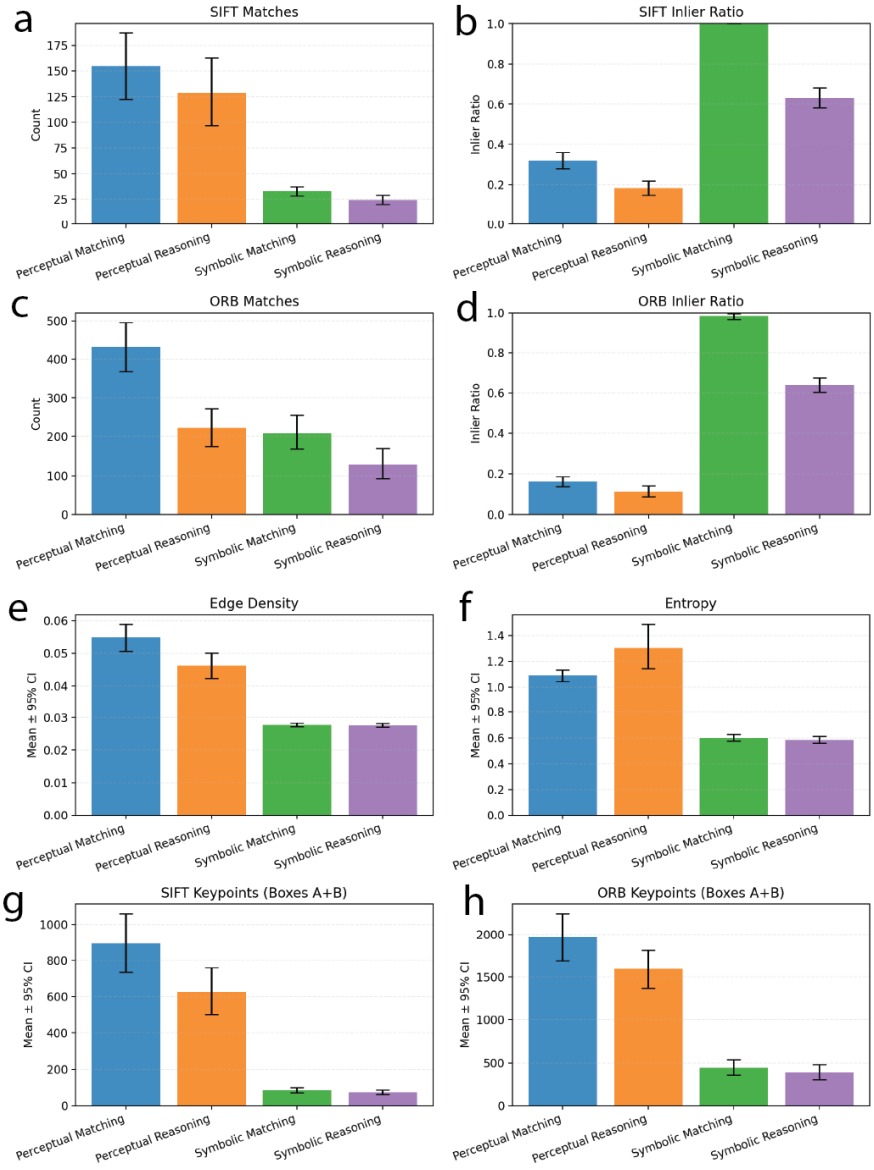
Supplements



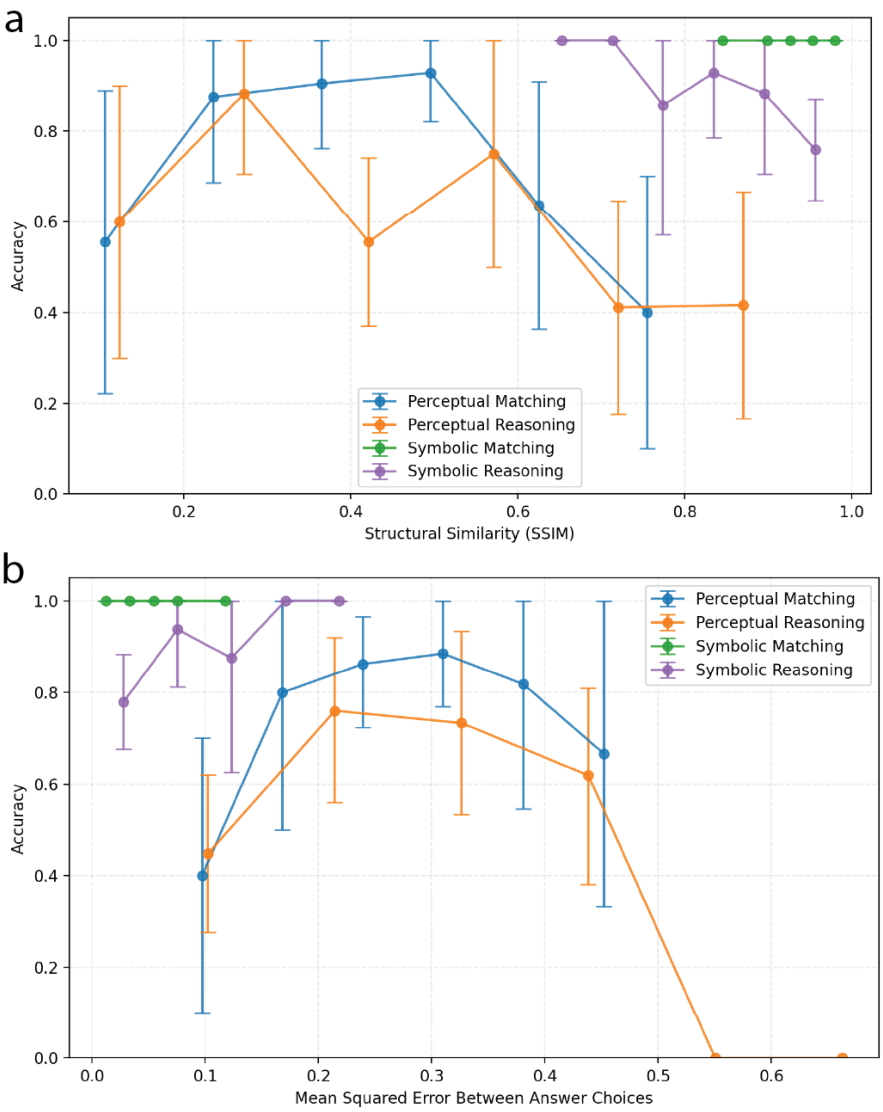
Supplemental Figure 1: Direct comparison of SIFT and ORB through the full model. a) Performance of the model on the Perceptual Matching condition using ORB, SIFT and hybrid ORB-SIFT features. b) Performance of the model on the Perceptual Reasoning condition using ORB, SIFT and hybrid ORB-SIFT features. c) Performance of the model on the Symbolic Matching condition using ORB, SIFT and hybrid ORB-SIFT features. d) Performance of the model on the Symbolic Reasoning condition using ORB, SIFT and hybrid ORB-SIFT features.



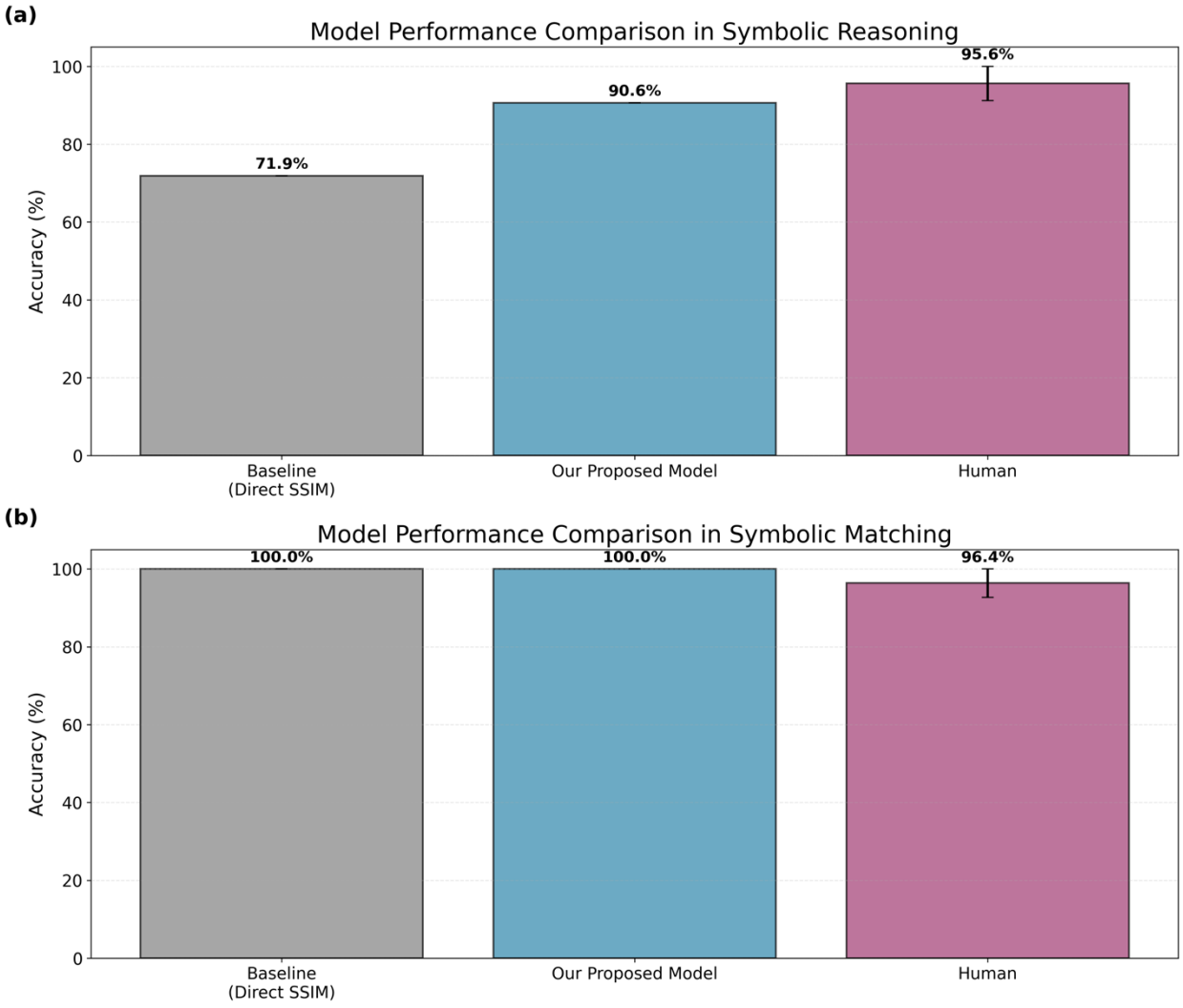
Supplemental Figure 2: Sensitivity analysis of threshold parameters and their effect on model performance. The heatmap shows model accuracy (color-coded from dark blue = low to yellow = high) as a function of two active threshold parameters, Threshold 1 and Threshold 2. Each parameter was systematically varied from 0.10 to 0.90 in 17 increments, resulting in 289 total combinations (17 × 17 grid). Each configuration was tested across 96 trials in the symbolic reasoning condition. Overall accuracy ranged from 77.1% to 93.8% (mean = 84.5% ± 3.7%), revealing a 16.7-point performance spread across the parameter space. The optimal configuration occurred at Threshold 1 = 0.85 and Threshold 2 = 0.75, achieving 93.8% accuracy.



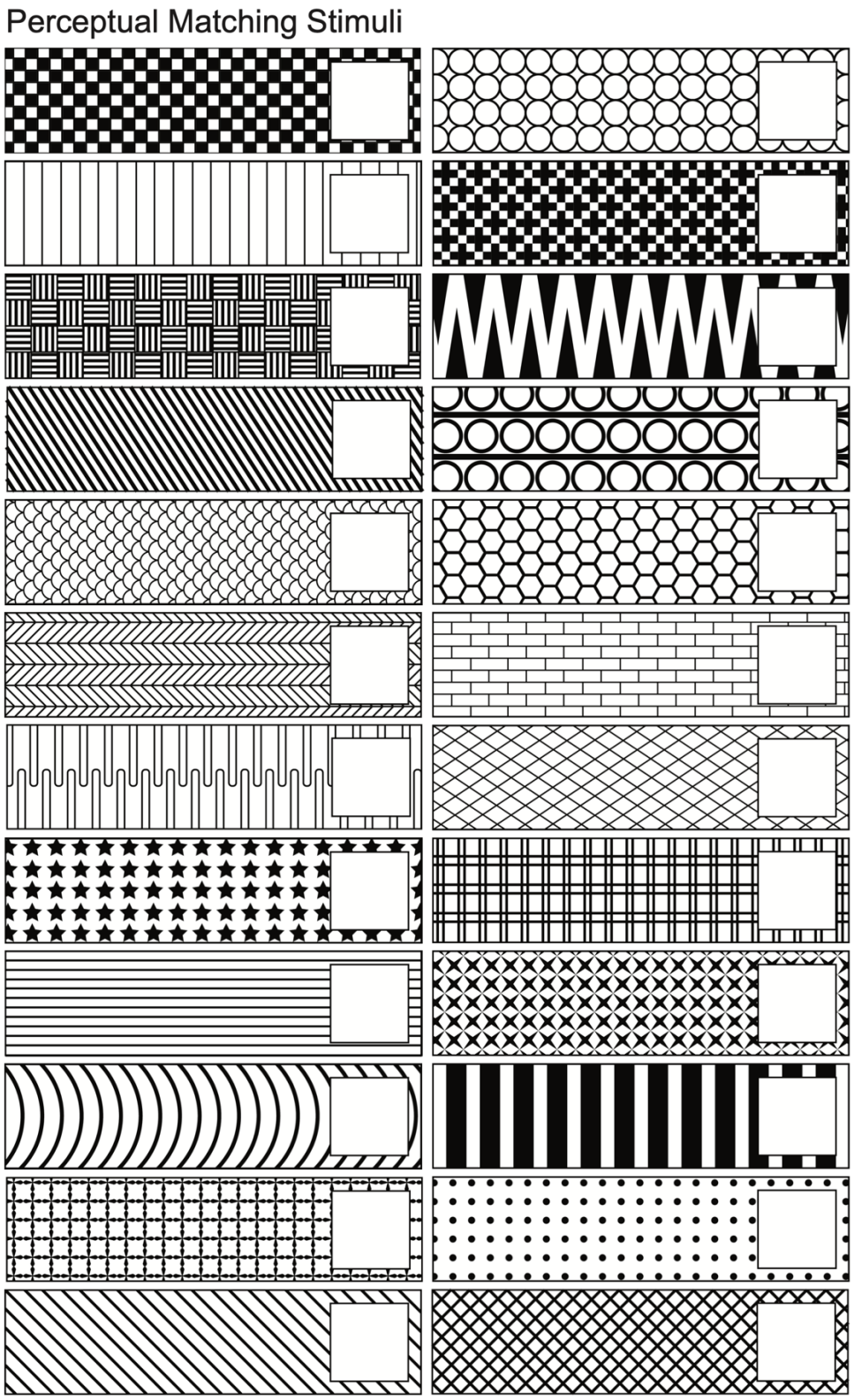
Supplemental Figure 3: Task difficulty and visual complexity metrics across four task conditions. a–d) Bar charts show mean ± 95% bootstrap confidence intervals (2,000 resamples; fixed RNG seed) for **SIFT Matches, SIFT Inlier Ratio, ORB Matches, and ORB Inlier Ratio** across the four conditions (*Perceptual Matching, Perceptual Reasoning, Symbolic Matching, Symbolic Reasoning*). "Matches" denotes the number of descriptor correspondences; **Inlier Ratio** represents the fraction of matches deemed geometrically consistent by RANSAC during similarity-transform estimation (bounded 0–1; y-axis limited accordingly). **e–h)** Bar charts show mean ± 95% bootstrap confidence intervals (2,000 resamples; fixed RNG seed) for **visual-complexity metrics** computed on the cue images: **Edge Density, Shannon Entropy,** and feature counts **(SIFT Keypoints** and **ORB Keypoints)** aggregated over Boxes A + B. Error bars reflect percentile confidence intervals; x-axis labels denote the four task conditions (*Perceptual Matching, Perceptual Reasoning, Symbolic Matching, Symbolic Reasoning*).



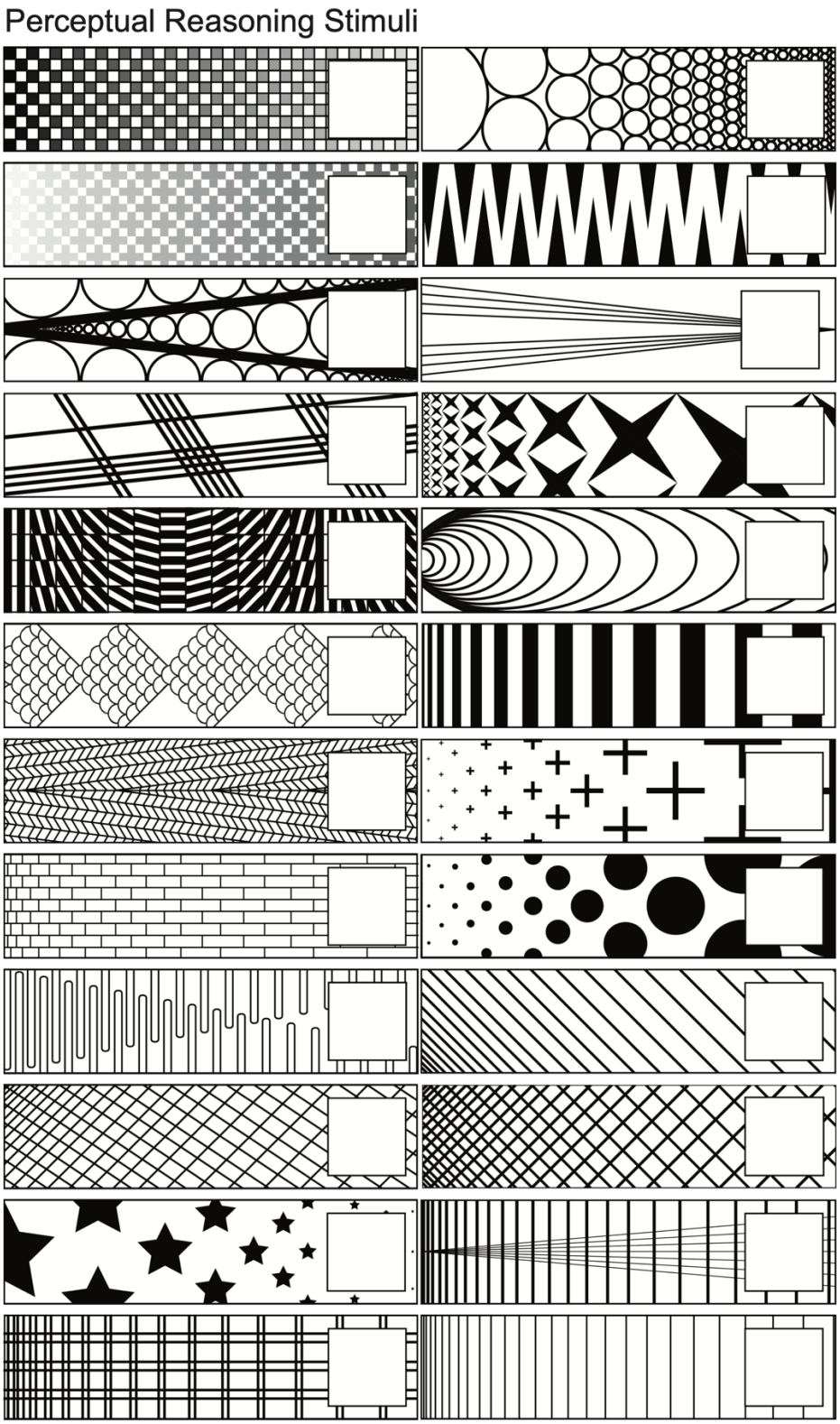
Supplemental Figure 4. Model accuracy as a function of answer-choice similarity. a) Model accuracy (y-axis) plotted against the **structural similarity index (SSIM)** between the two answer choices (x-axis). SSIM ranges from 0 (dissimilar) to 1 (identical) and reflects perceptual similarity. **b)** Model accuracy (y-axis) plotted against the **mean squared error (MSE)** between the two answer choices (x-axis). MSE represents the normalized per-pixel difference between candidate answers; lower values indicate higher similarity. For both panels, points show mean accuracies within uniformly spaced bins (e.g., 20 bins), connected to emphasize trends. Shaded regions denote **95% bootstrap confidence intervals** (2,000 resamples) around the mean accuracy in each bin.



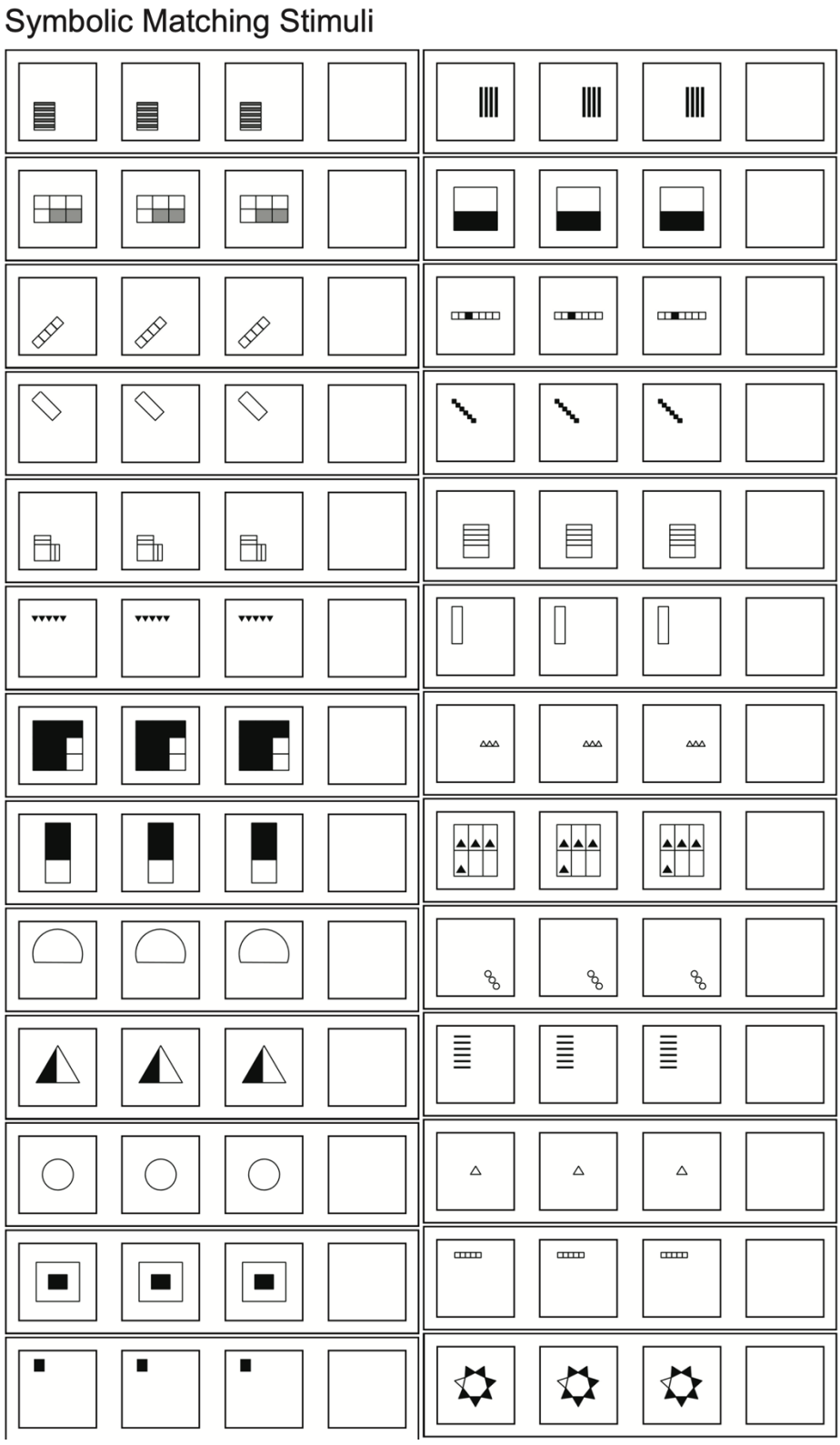
Supplemental Figure 5: Comparison with a template matching strategy. Our Proposed Model substantially outperforms a direct SSIM baseline on symbolic reasoning and closely approaches human performance. Top row (Symbolic Reasoning): Baseline achieves 71.9%, Our Proposed Model reaches 90.6%, and Humans achieve 95.6% (error bars denote 95% confidence interval, capped within 0–100%). Bottom row (Symbolic Matching): Baseline and Our Proposed Model both achieve 100%, while Humans achieve 96.36% (±18.73%, capped). Bars show mean accuracies; human error bars reflect inter-subject variability.



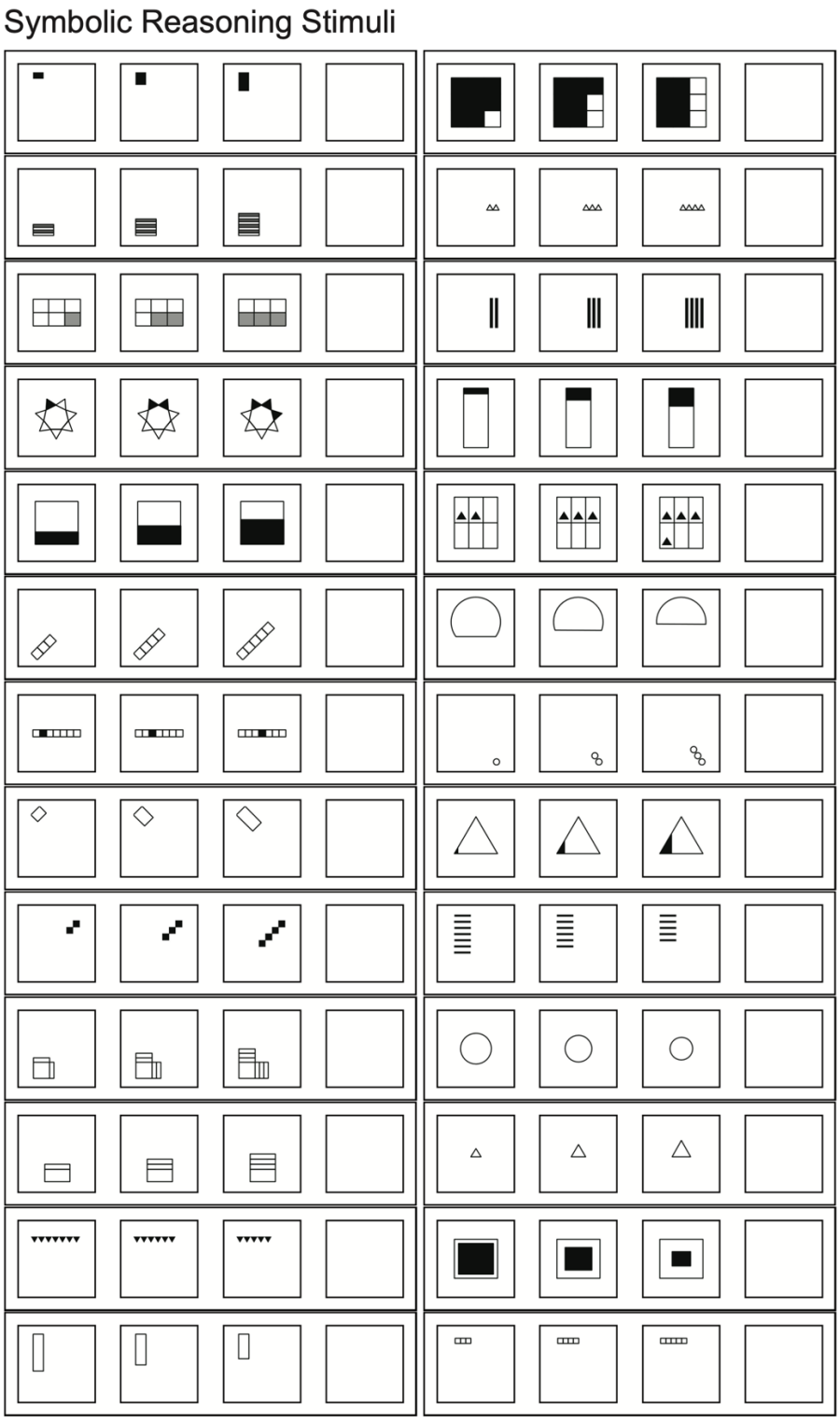
Supplemental Figure 6: Additional Examples for Perceptual Matching Stimuli.



Supplemental Figure 7: Additional Examples for Perceptual Reasoning Stimuli.



Supplemental Figure 8: Additional Examples for Symbolic Matching Stimuli.



Supplemental Figure 9: Additional Examples for Symbolic Reasoning Stimuli.