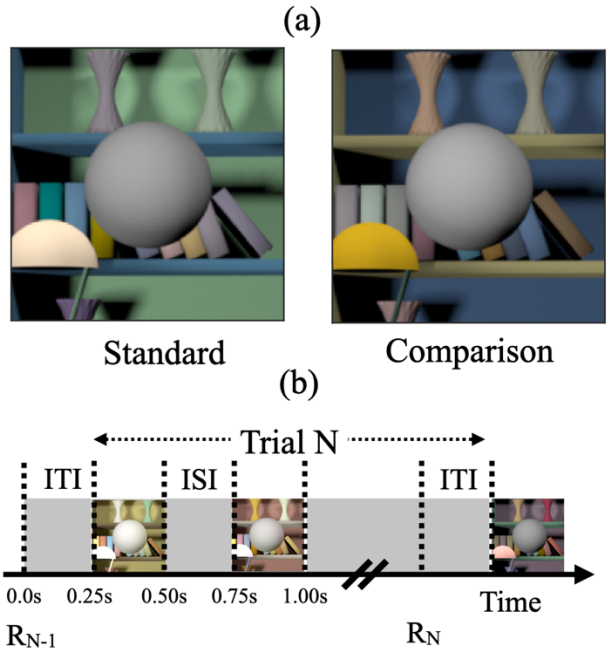

Figure 1: (a) Psychophysical task. (Adapted from Figure 1 in Singh, Burge, & Brainard, 2022) The psychophysical task involved comparing two images, a standard image and a comparison image, on each trial and selecting the image with the lighter target object. The target object was an achromatic sphere at the center of the image. The images were generated computationally by graphically rendering models of 3D scenes. They were displayed on a color-calibrated monitor. This panel shows examples of standard and comparison images. The reflectance spectrum of the target object was spectrally flat, and the target object appeared gray. The reflectance of the target object in the standard image was held fixed and it changed for the comparison image. In this panel, the target object in the comparison image is lighter. We measured the fraction of times the observers chose the target object in the comparison image to be lighter as a function of the lightness of the target object in the comparison image. The proportion comparison chosen data was used to determine the lightness discrimination threshold (Figure 2). We studied how the lightness discrimination thresholds changed as the trial-to-trial variability in the reflectance spectra of the background objects and the intensity of the light sources increased. **(b) Trial sequence:** R_{N-1} indicates the recording of the observer's response for the $(N-1)^{th}$ trial. The N^{th} trial begins 250ms after the completion of the $(N-1)^{th}$ trial (Inter Trial Interval, ITI = 250ms). In the N^{th} trial, the standard and comparison images are presented for 250ms each with a 250ms inter-stimulus interval (ISI) in between the two images. The order of the standard and comparison images is chosen in pseudorandom order. The observer records their choice by pressing a button on a gamepad after both images have been presented and removed from the screen. The observers could take as long as they wish before making their choice. The recording of their choice is indicated by R_N in the panel. The next trial begins 250ms after the choice has been recorded.

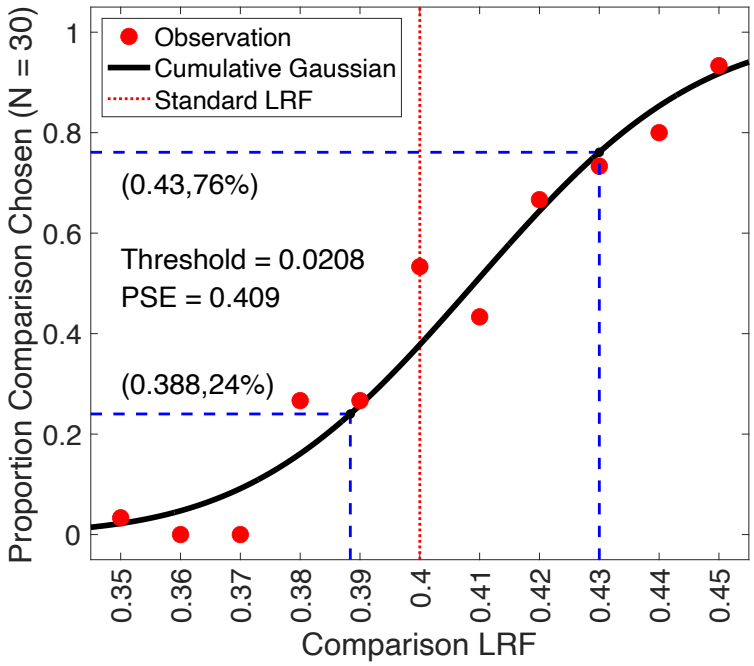

Figure 2: Psychometric function: We measured the proportion of times the observers selected the target in the comparison image to be lighter as a function of the LRF (lightness reflectance factor) of the target object. We collected 30 responses for each of the 11 equally spaced values of the comparison image target object LRF, ranging from 0.35 to 0.45. The LRF of the target object in the standard image was 0.40. The LRF of the target object in the comparison image was selected in a pseudorandom order. To analyze the data, we used maximum likelihood methods to fit a cumulative normal function to the proportion-comparison-chosen data. We imposed constraints on the guess rate and lapse rate, requiring them to be equal and within the range of 0 to 0.05. The threshold was determined as the difference between the LRF values corresponding to a proportion-comparison-chosen of 0.76 and 0.50, obtained from the cumulative normal fit. The figure presented here illustrates the data for observer 0003 in the second block of the *background reflectance variation* experiment for the no-variation ($\sigma^2 = 0.00, \delta = 0.00$) condition. The discrimination threshold was measured to be 0.0208. The point of subjective equality (PSE), which corresponds to a proportion of 0.5 in the comparison task, was found to be 0.409. The lapse rate for this particular fit was determined to be 0.00.

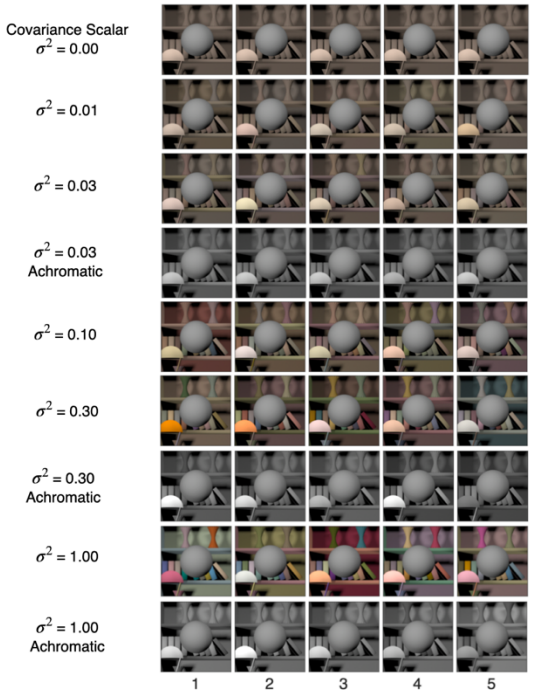

Figure 3: Background reflectance variation: We studied two types of variations in the reflectance spectra of background objects in the scene: chromatic variation and achromatic variation. In chromatic variation, the reflectance spectra could take any shape, and the objects varied in their luminance and chromaticity. In the achromatic variation, the reflectance spectra were spectrally flat, and the objects appeared gray and varied only in their luminance. The spectra were chosen from a multivariate normal distribution that modeled the statistics of natural reflectance spectra. The covariance matrix of the multivariate normal distribution was multiplied by a scalar to control the variance in the samples. We generated images at six logarithmically spaced values of the covariance scalar for chromatic variation and at three values of the covariance scalar for achromatic variations. The figure shows five typical images for each of these nine conditions. For each condition, we generated 1100 images, 100 images at 11 linearly spaced values of target object LRF in the range [0.35, 0.45]. The target object in each image in the figure is at LRF = 0.4.

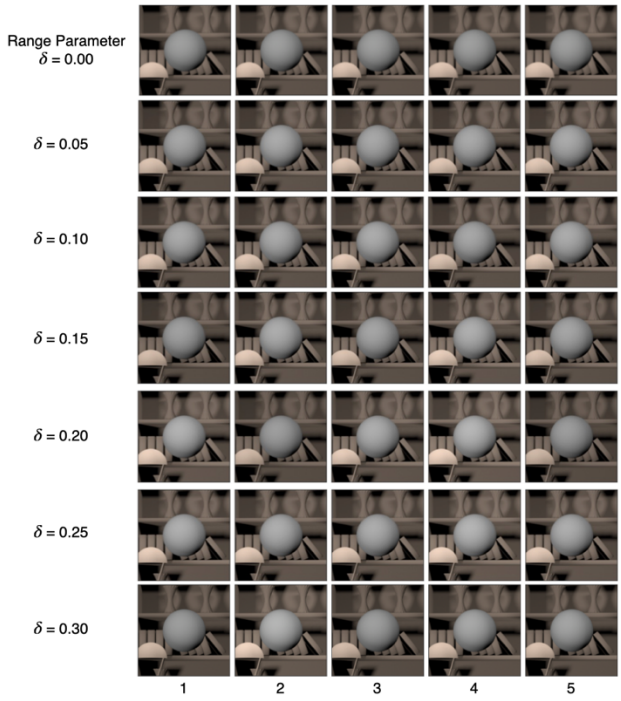

Figure 4: Light intensity variation: The shape of the power spectrum of the light sources in the scene was chosen to be CIE reference illuminant D65. The intensity of the power spectrum was varied by multiplying the normalized D65 spectrum with a scalar sampled from a log uniform distribution in the range [1- δ, 1+ δ]. The amount of variation was controlled by changing the value of the range parameter δ. We generated images at seven linearly spaced values of the range parameter in the range [0.00, 0.30]. For each value of the range parameter, we generated 1100 images, 100 images at each value of the target object LRF in the range [0.35, 0.45]. The figure shows five sample images at each of the seven values of the range parameter. The target object in each image in the figure has the same LRF of 0.40.

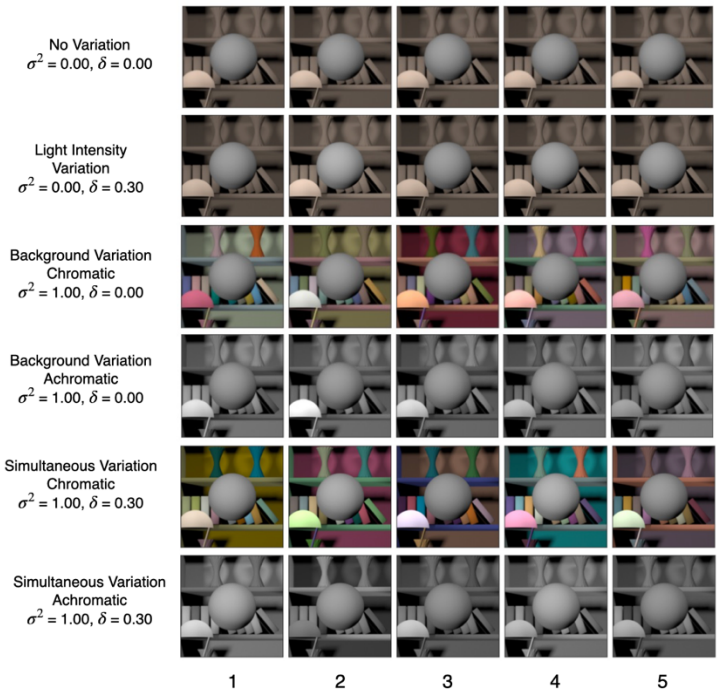

Figure 5: Simultaneous variation: This figure shows five sample images for the six conditions studied in *simultaneous variation* experiment. We generated 1100 images for each of these conditions, 100 images at each value of the target object LRF in the range [0.35, 0.45].

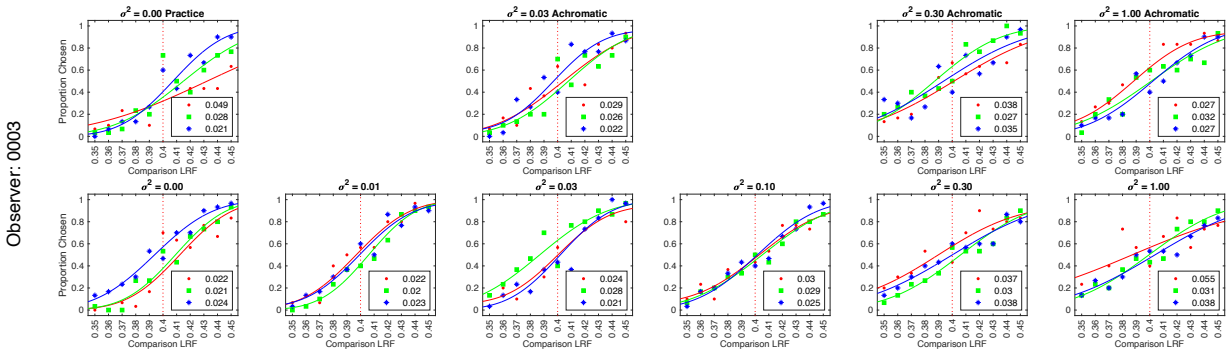

Figure 6: Psychometric functions of observer 0003 for *background reflectance variation* experiment: We measured the proportion-comparison-chosen data for the nine conditions separately in three blocks for each observer. The figure shows the psychometric function for observer 0003. The psychometric functions for all six observers are shown in Figure S2. A cumulative normal function was fit to the data from each block to determine the discrimination threshold (see Figure 2). The legend provides the estimated lightness discrimination threshold for each block, obtained from the cumulative fit. The first panel in the top row shows the data and thresholds for the practice session. An observer was selected for the experiment only if the average of their last two discrimination threshold measurements in the practice session was less than 0.30. The last three panels in the top row show the data for the three achromatic conditions. The bottom row shows the data for the chromatic variation conditions.

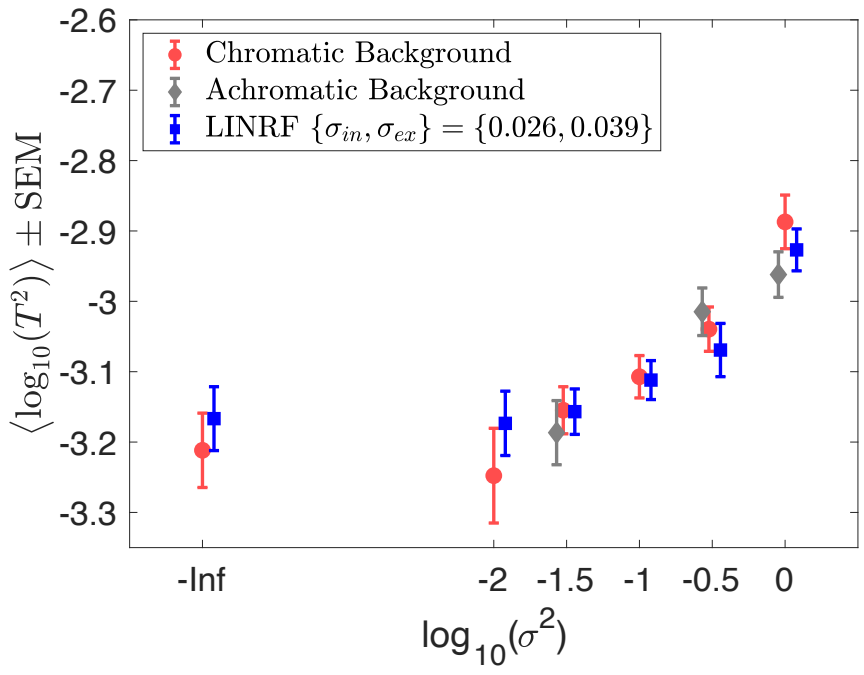

Figure 7: Background reflectance variation increases lightness discrimination thresholds. Mean (N = 6) log squared threshold vs log covariance scalar from human psychophysics for chromatic (red circles) and achromatic conditions (gray diamonds). The error bars represent +/- 1 SEM taken between observers. The threshold of the linear receptive field (LINRF) model was estimated by simulation for the six values of the covariance scalar (blue squares). The blue error bars show +/- 1 standard deviation estimated over 10 independent estimates of the LINRF model parameters. The legend shows the parameters of the linear receptive field (LINRF) model fit. The data has been jittered for ease of viewing. A comparison of the thresholds with the previously published data in Singh, Burge, Brainard 2022 is shown in Figure S3.

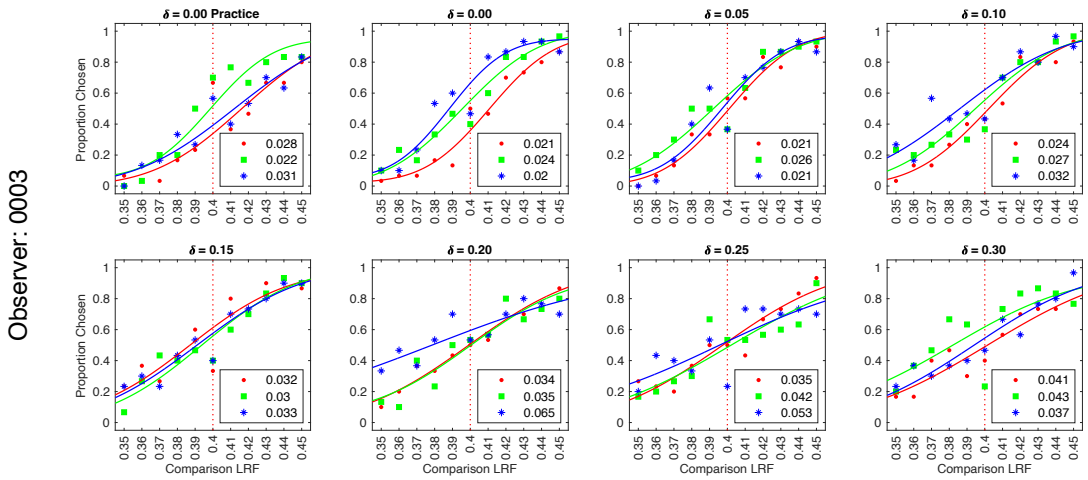

Figure 8: Psychometric functions for observer 0003 for light source intensity variation experiment: Same as Figure 6, but for the *light source intensity variation* experiment. The figure shows the proportion comparison chosen data for the selection session and the seven conditions for observer 0003. The psychometric functions for all observers are shown in Figure S4.

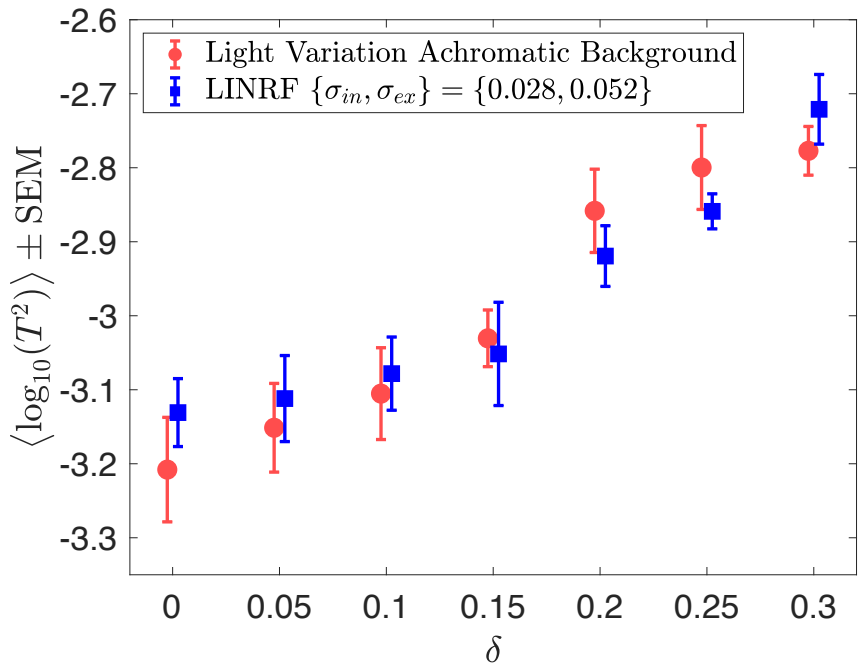

Figure 9: Light source intensity variation increases lightness discrimination threshold. Mean (N = 5) log squared threshold vs range parameter from human psychophysics for the seven *light source intensity variation* conditions (red circles). The error bars represent +/- 1 SEM taken between observers. The threshold of the linear receptive field (LINRF) model was estimated by simulation for the seven values of the range parameters (blue squares). The blue error bars show +/- 1 standard deviation estimated over 10 independent estimates of the LINRF model parameters. The legend shows the parameters of the LINRF model fit. The data has been jittered for ease of viewing. The data for all six observers is shown in Figure S5.

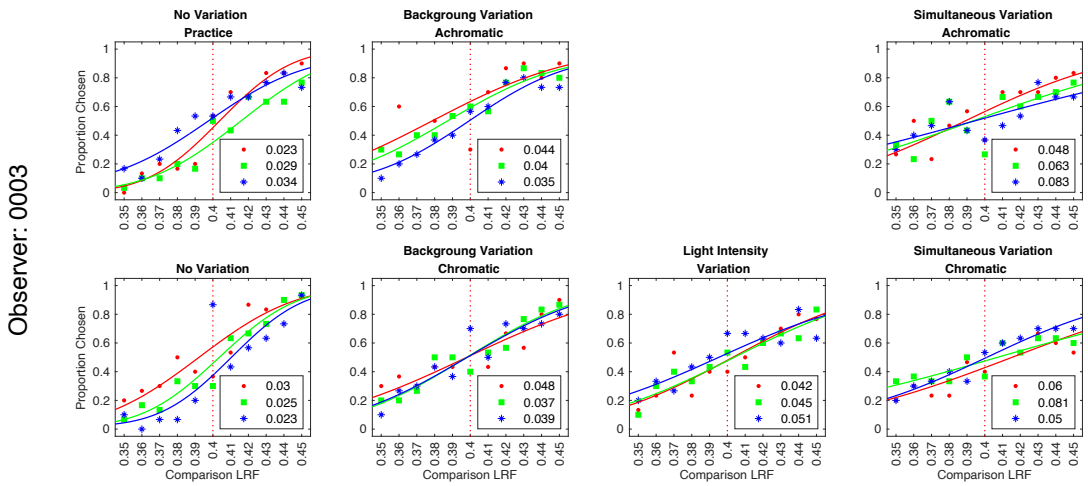

Figure 10: Psychometric functions for observer 0003 for simultaneous variation experiment: Same as Figures 6 and 8, but for simultaneous variation experiment. The figure shows the proportion comparison chosen data for the selection session and the six conditions for observer 0003. The data for all observers are shown in Figure S6.

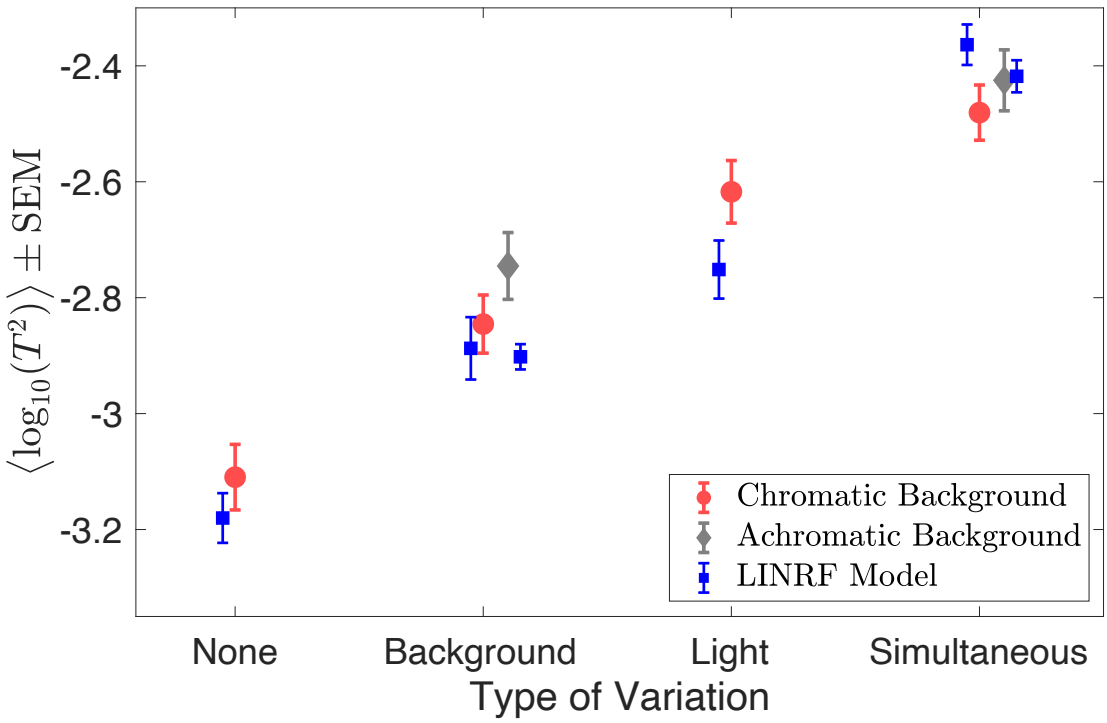

Figure 11: Discrimination thresholds for simultaneous variation of two sources are higher than individual discrimination thresholds. Mean (N = 6) log squared threshold for the six conditions in simultaneous variation experiment. The error bars represent +/- 1 SEM taken between observers. The data for chromatic (red circles) and achromatic (gray diamonds) conditions have been plotted next to each other for visual comparison. The thresholds of the linear receptive field (LINRF) model (blue squares) were estimated using the parameters of the background variation condition (Figure 7) for the None, Background variation, and Simultaneous variation conditions and using the parameters of the *light intensity variation* experiment (Figure 9) for the Light condition. The blue error bars show +/- 1 standard deviation estimated over 10 independent estimates of the LINRF model parameters. See Figure S7 for LINRF model thresholds with the same set of parameters for all conditions.

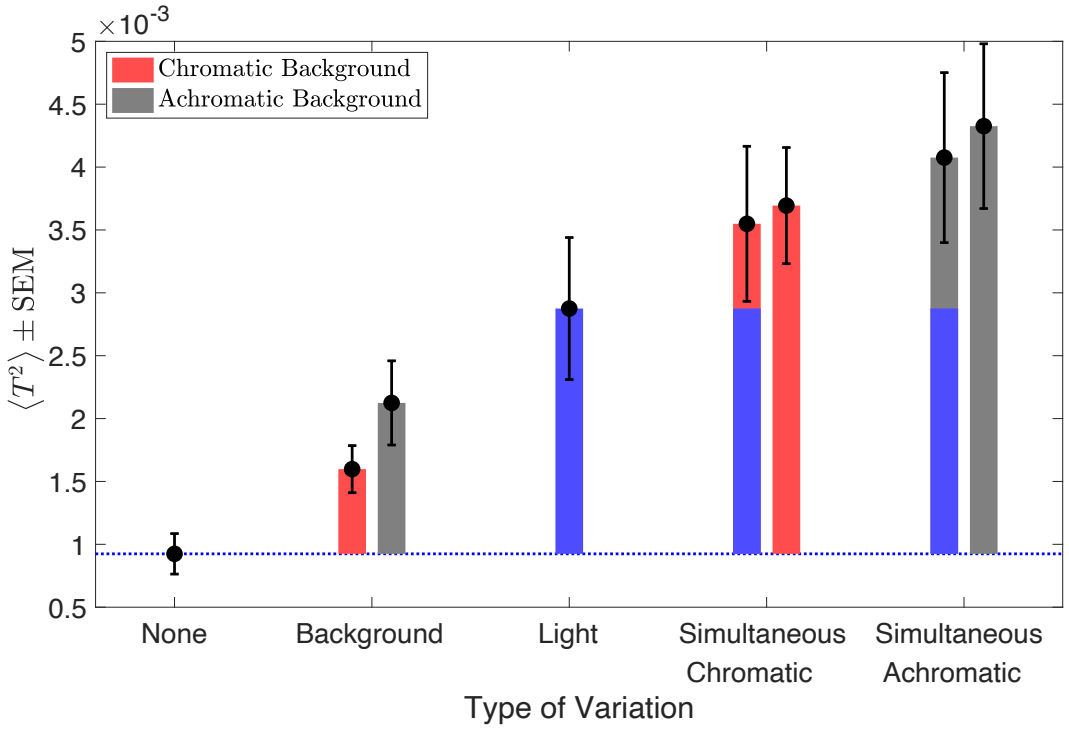

Figure 12: Extrinsic noise of independent variations adds linearly for simultaneous variation: Mean squared thresholds (N=6) for the six conditions in simultaneous variation experiment (black circles). The black error bars represent +/- 1 SEM taken between observers. The bars (red, gray, blue) represent the increase in squared thresholds compared to the no-variation condition (blue dotted line). For the simultaneous variation conditions, the bars on the right (bars with one color, red or gray) represent the increase in the measured squared threshold and the bars on the left (stacked bars of two different colors) represent the increase in the sum of the squared threshold of the light intensity variation (blue bar) and the corresponding background variation conditions (red or gray).

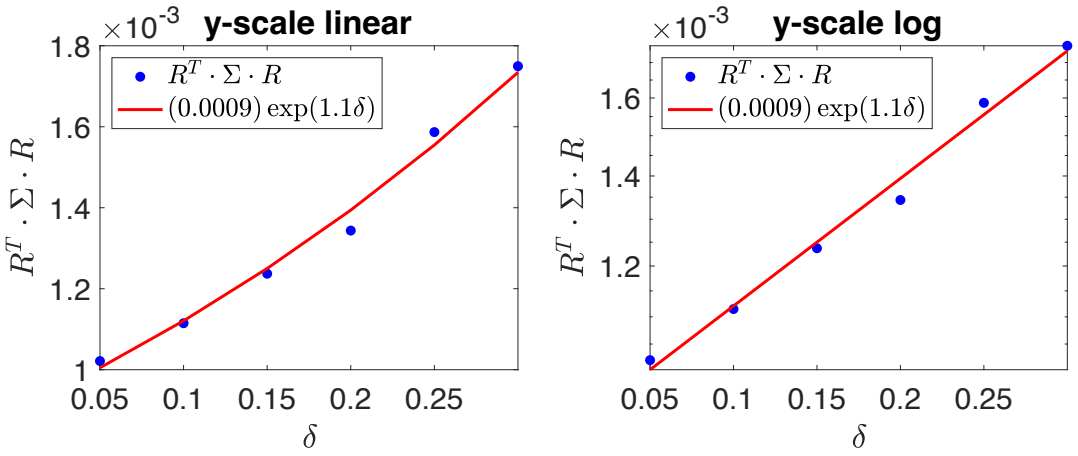

Figure S1: Estimation of extrinsic noise for *light source intensity variation* experiment: Plot of the variance ($R^T \Sigma R$) as a function of the range parameter δ on a linear (left panel) and logarithmic (right panel) scale. We fit the function with an exponential of the form $A * \exp(B \cdot \delta)$. The variance in the extrinsic noise is estimated as the value of the fit at $\delta = 1$.

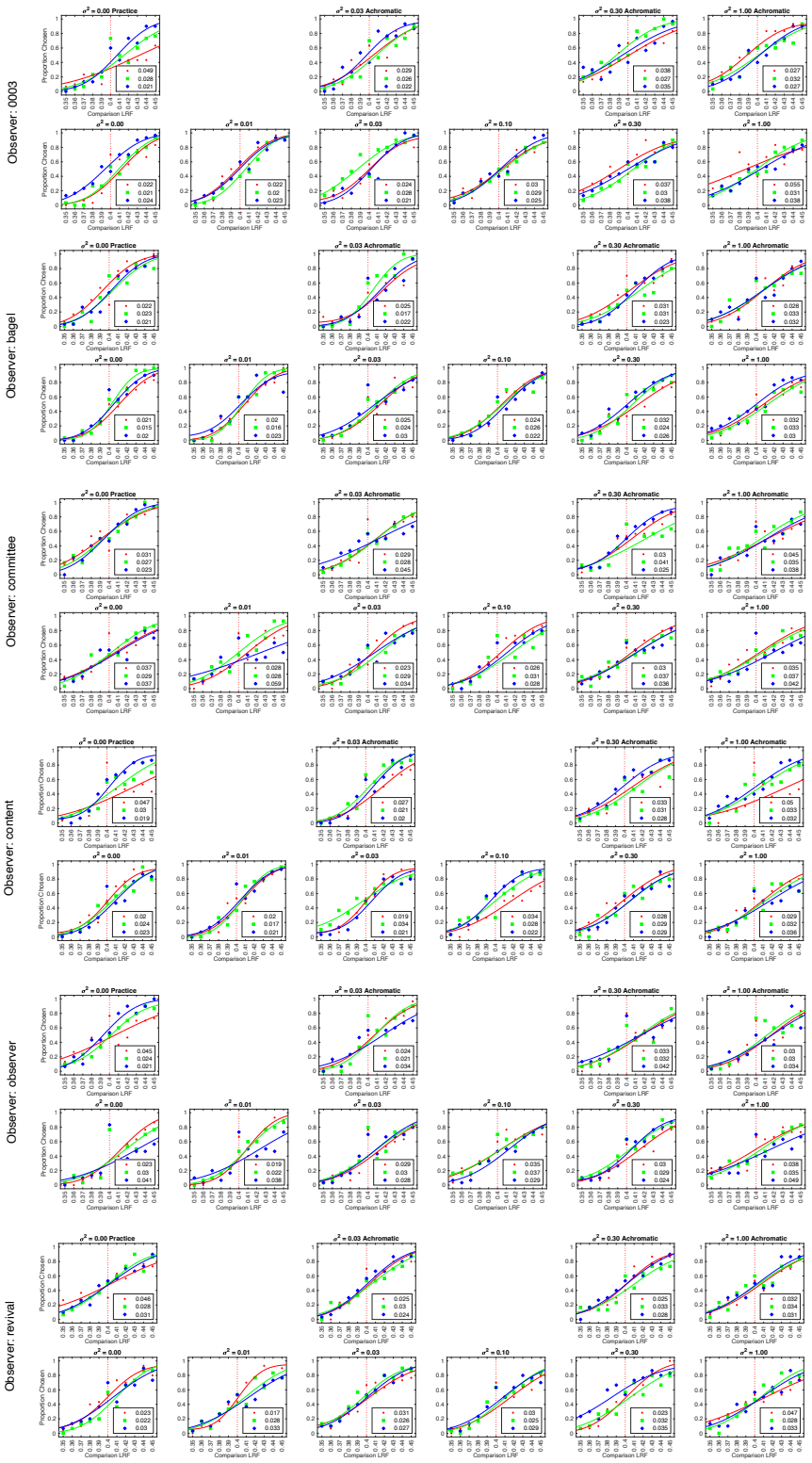

Figure S2: Psychometric functions for all observers for background reflectance variation experiment. Same as Figure 6, for all observers retained in the *background reflectance variation* experiment.

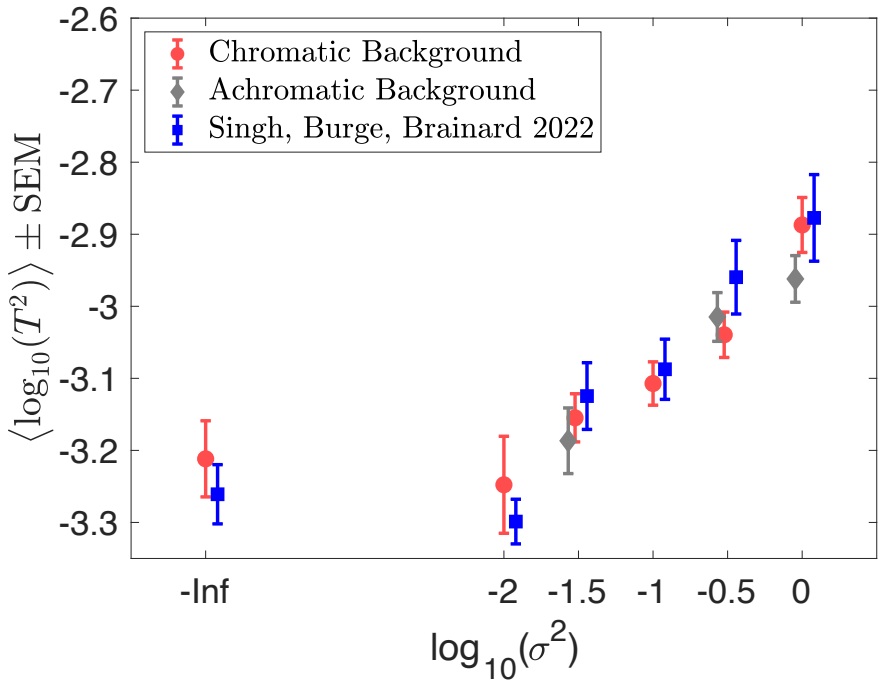

Figure S3: Comparison with Singh, Burge, Brainard 2022. Lightness discrimination thresholds for background variation condition measured in *background reflectance variation* experiment and previously reported data from Singh, Burge, Brainard (2022). The previous experiment only had chromatic conditions and made three threshold measurements for each condition for 4 naïve observers. In this work, *background reflectance variation* experiment had both chromatic and achromatic conditions and measured thresholds for six observers. The experiments were otherwise the same.

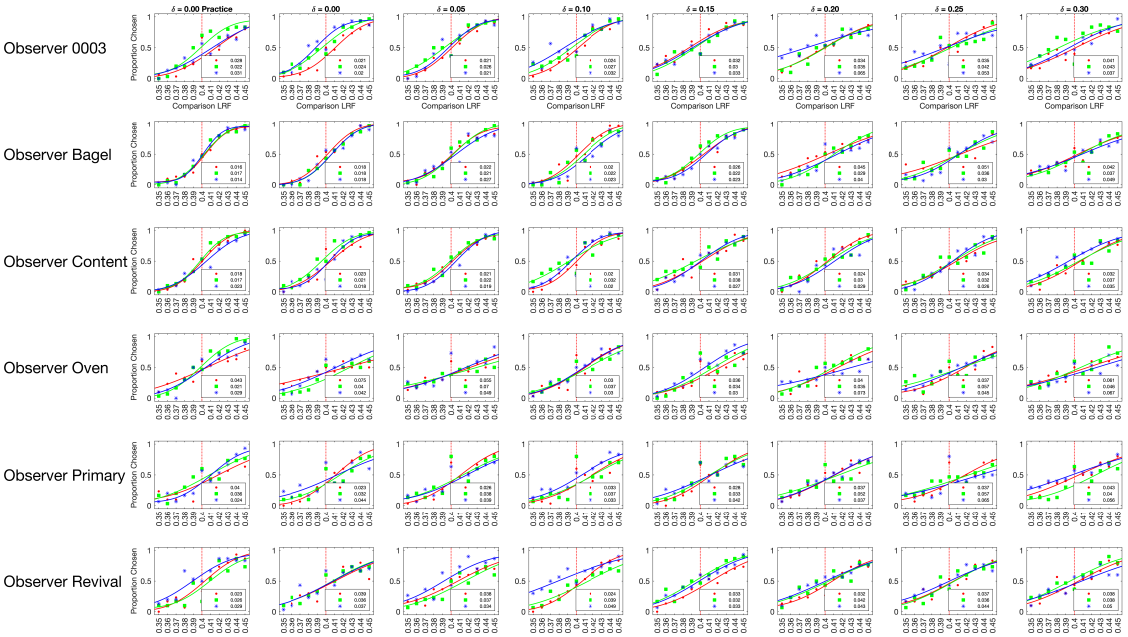

Figure S4: Psychometric functions for all observers for light intensity variation experiment. Same as Figure 8, for all observers retained in the *light source intensity variation* experiment.

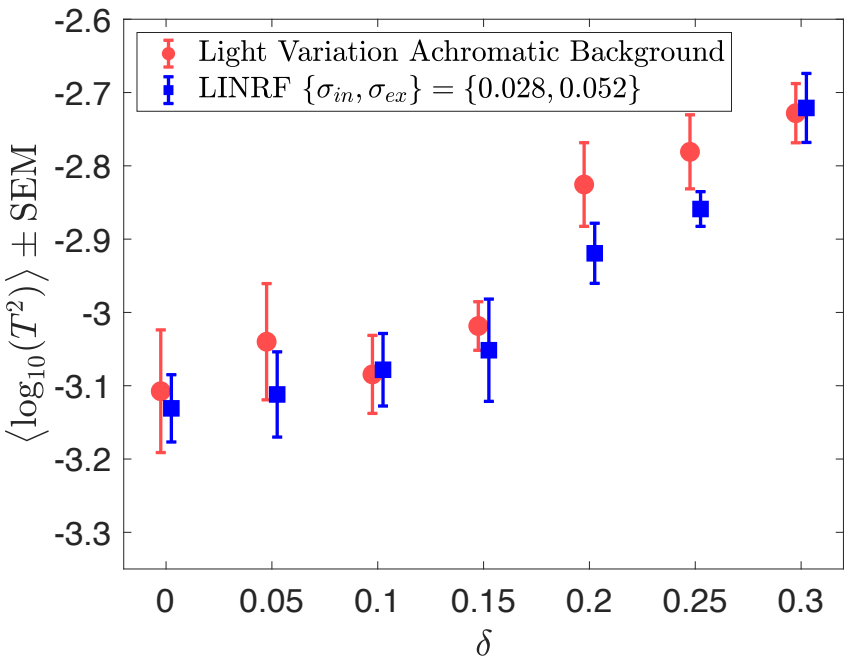

Figure S5: Same as Figure 9, for all six observers retained in the *light source intensity variation* experiment. The parameters for the LINRF model are the same as in Figure 9.

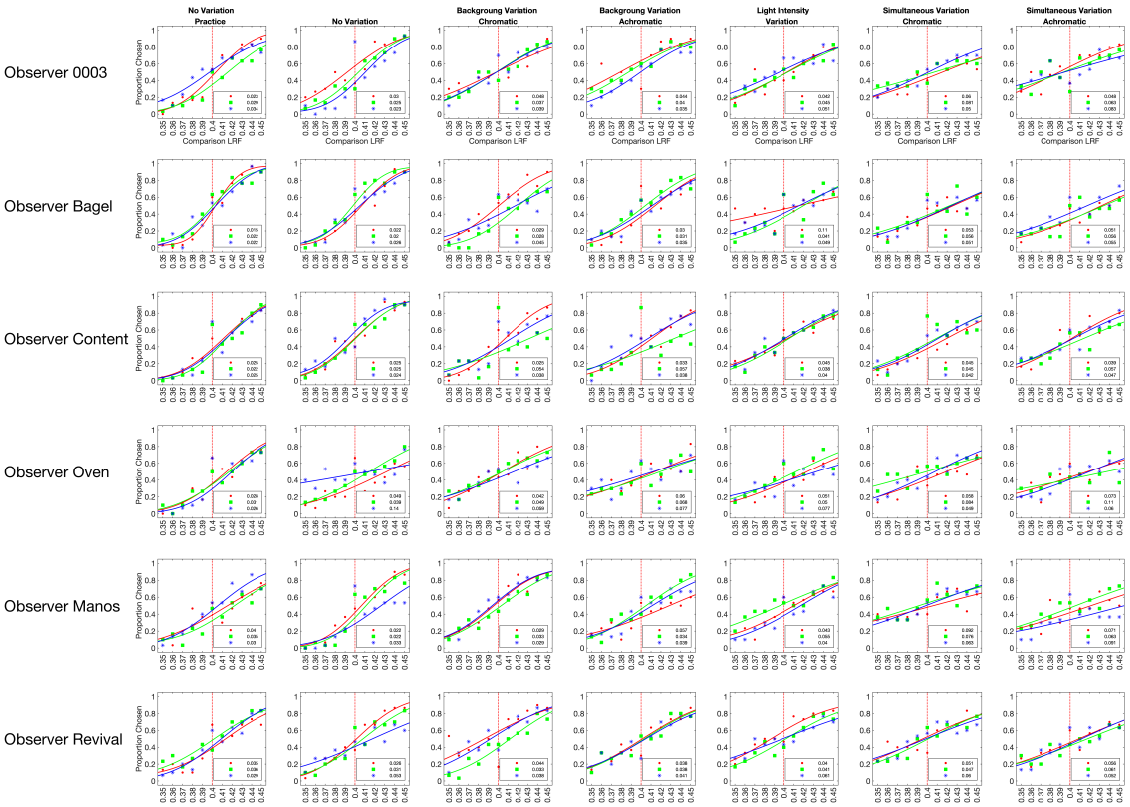

Figure S6: Psychometric functions for all observers for Simultaneous variation experiment. Similar to Figure 10, for all observers retained in the *simultaneous variation* experiment.

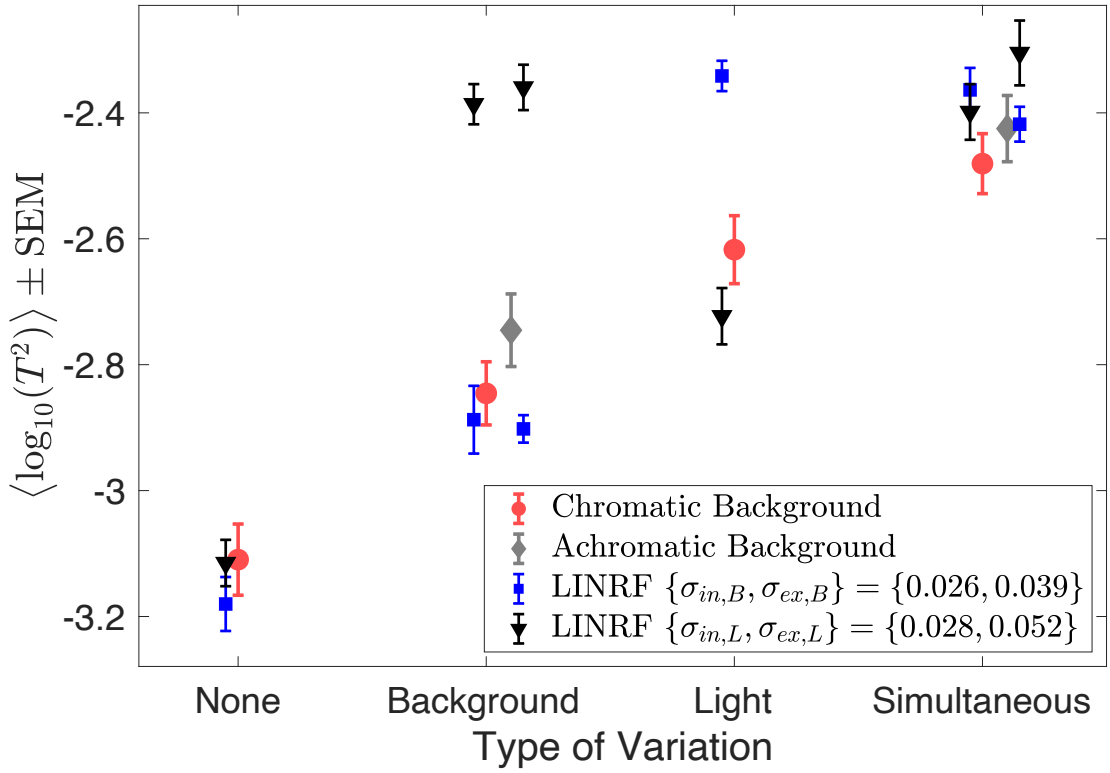

Figure S7: Same as Figure 11, but the thresholds of the linear receptive field (LINRF) model were estimated using the same set of parameters for all six conditions studied in Simultaneous variation experiment. Blue square markers show log squared thresholds estimated using the parameters of the Background reflectance variation experiment (Figure 7). Black triangular markers show log-squared thresholds estimated using the parameters of the Light intensity variation condition (Figure 9). The blue error bars show +/- 1 standard deviation estimated over 10 independent estimates of the LINRF model parameters. The parameters of the *background reflectance variation* condition (blue squares) predict the thresholds of the no-variation condition, the background reflectance variation condition, and the simultaneous variation condition quite well, but fail to predict the threshold of the light source intensity variation condition. Similarly, the parameters of the *light source intensity variation* experiment (black triangles) predict the thresholds of the no-variation condition, the light source intensity variation condition, and the simultaneous variation condition quite well, but fail to predict the threshold of the background variation condition. This could possibly be because the observers in the three experiments were different. Future work would aim at studying these conditions using the same set of observers.