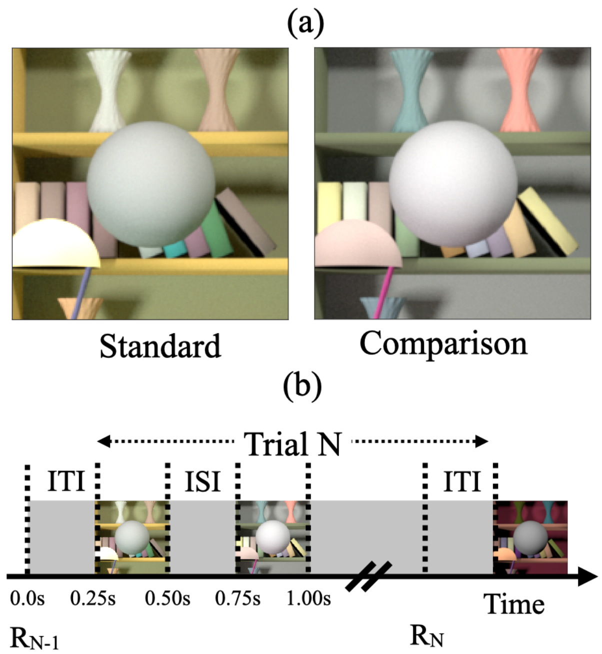

Figure 1: (a) Psychophysical task. On every trial of the experiment, human observers viewed two images, a standard image and a comparison image, and indicated the image in which the spherical target object at the center of the image was lighter. The images were computer graphics renderings of 3D scenes. 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. Fraction comparison chosen data was used to determine 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 $(N\text{-}1)^{th}$ trial. The N^{th} trial begins 250ms after the completion of the $(N\text{-}1)^{th}$ 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.

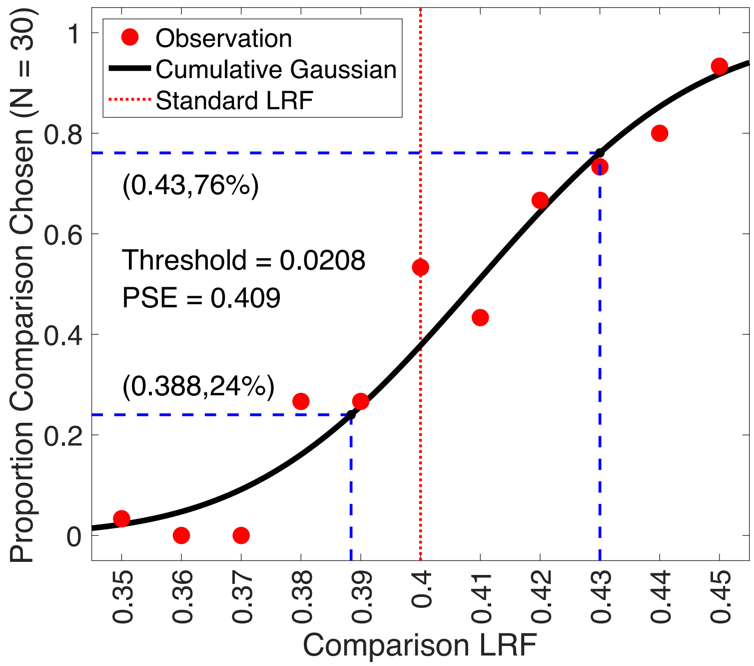

Figure 2: Psychometric function: We recorded the proportion of times the observers chose the target in the comparison image to be lighter as a function of the LRF of the target object in the comparison image. We collected 30 responses each at 11 linearly spaced values of the comparison image target object LRF in the range [0.35, 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 chosen in a pseudorandom order. The proportion comparison chosen data was fit by a cumulative normal distribution using maximum likelihood methods. The guess rate and lapse rate were constrained to be equal and restricted to be in the range [0, 0.05]. The threshold was measured as the difference between the LRF at proportion comparison chosen equal to 0.76 and 0.50 as obtained from the cumulative normal fit. This figure shows the data for observer 0003 in the second block of background reflectance variation experiment (preregistered Experiment 6) for the no variation ($\sigma^2 = 0.00$) condition. The discrimination threshold was 0.0208. The point of subjective equality (PSE, the LRF at which proportion comparison chosen is 0.5) was 0.409. The lapse rate for this fit was 0.00.

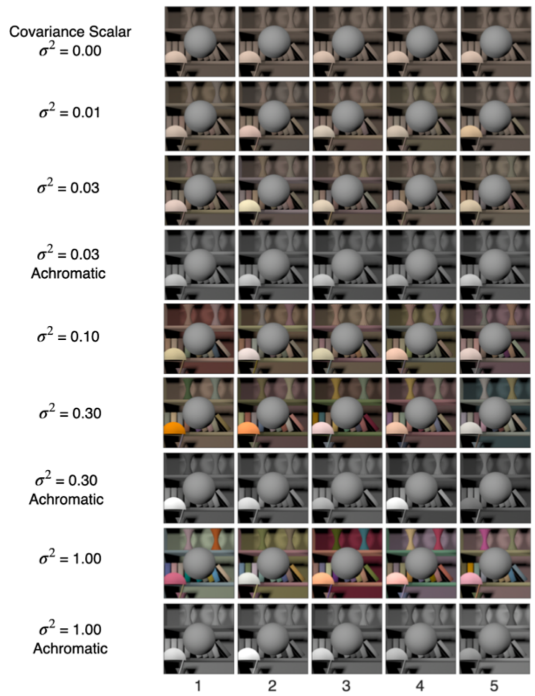

Figure 3: Background object reflectance variation: We studied two types of variations in the reflectance spectra of background objects in the scene: chromatic variation and achromatic variation. In chromatic variation, the reflectance spectra could take any shape, and the objects varied in their luminance and chromaticity. In 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 variation in the reflectance spectra was controlled by multiplying the covariance matrix of the distribution with a scalar. 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 value 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.

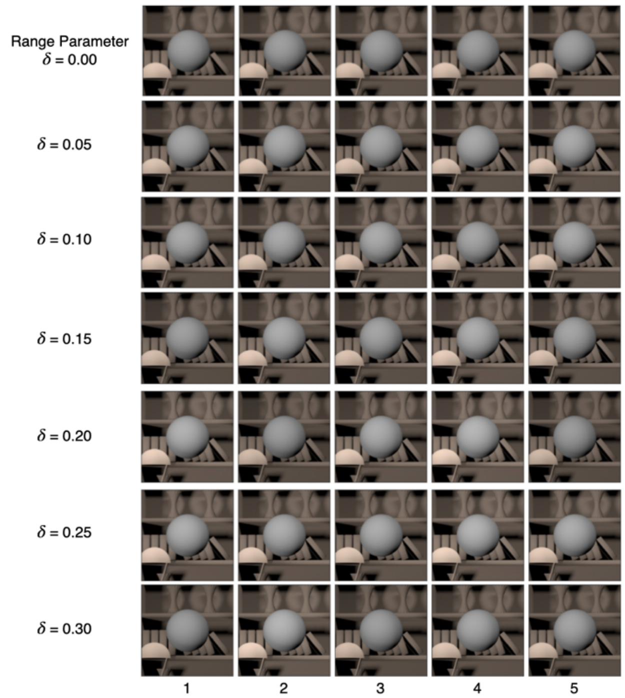

Figure 4: Light intensity variation: The shape of the power spectrum of the light sources in the scene was chosen to be CIE reference illuminant D65. The intensity of the power spectrum was varied by multiplying the normalized D65 spectrum with a scalar sampled from a log uniform distribution in the range $[1-\delta, 1+\delta]$. 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.

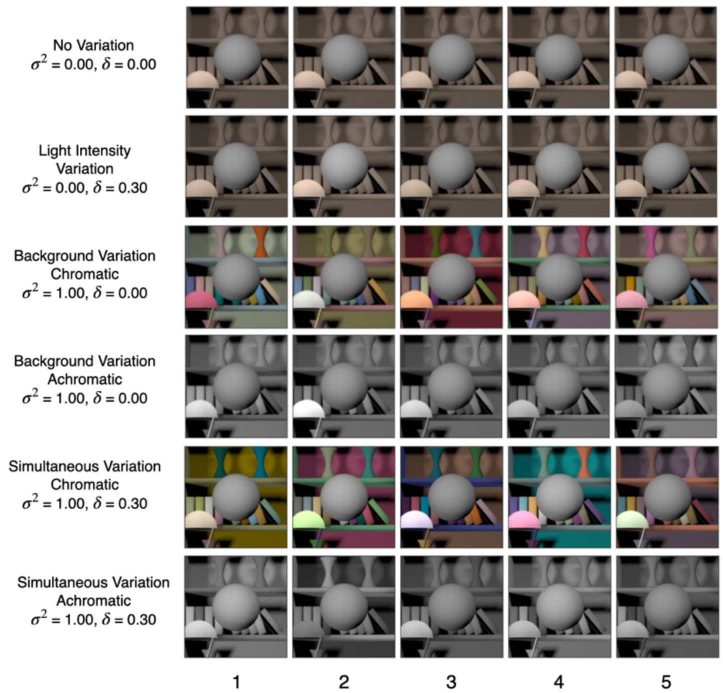

Figure 5: Simultaneous variation: This figure shows five sample images for the six conditions studied in preregistered experiment 8. We generated 1100 images for each of these conditions, 100 images at each value of the target object LRF in the range [0.35, 0.45].

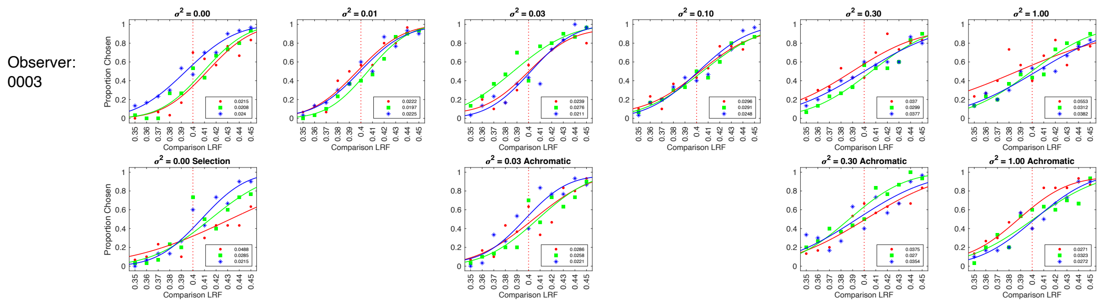

Figure 6: Psychometric functions for observer 0003 for background reflectance variation experiment: We measured the proportion comparison chosen data for the nine conditions separately in three blocks for each observer. The data for each block was fit with a cumulative normal to obtain the discrimination thresholds (see Figure 2). Each panel plots the measured values and the cumulative fit to the proportion comparison data for each of the three blocks, for observer 0003. The psychometric functions for all six observers are shown in Figure S2. The values in the legend provide the estimate of lightness discrimination threshold for each block obtained from the cumulative fit. The top row shows the data for chromatic variation conditions. The last three panels in the bottom row show the data for the three achromatic conditions. The first panel in the bottom row shows the data and thresholds for the selection session. The selection session was a practice session in which the thresholds for the no variation condition was measured three times. An observer was selected for the experiment only if the average of their last two discrimination threshold measurements in the selection session was less than 0.30.

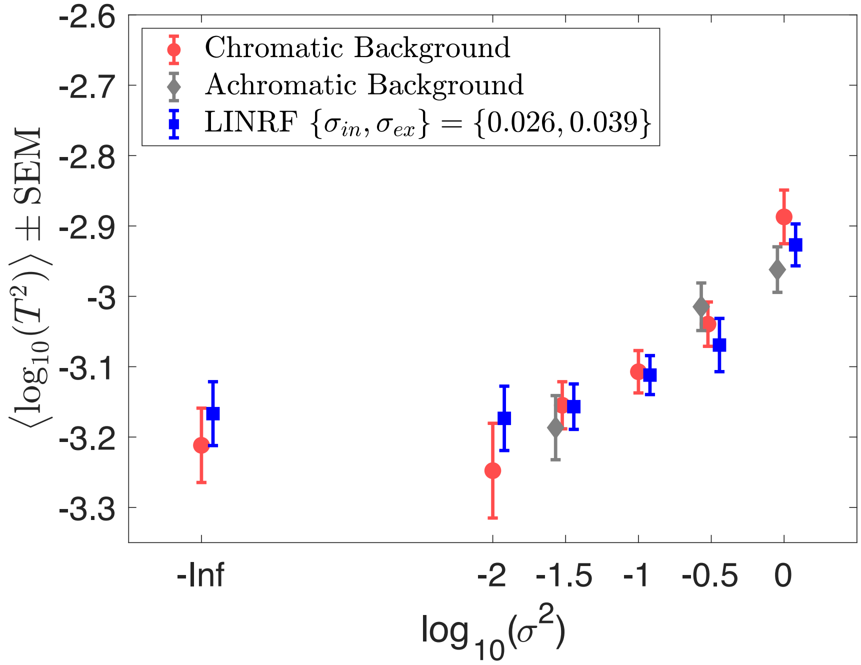

Figure 7: Background variation increases lightness discrimination thresholds. Mean (N = 6) log squared threshold vs log covariance scalar from human psychophysics for chromatic (red circles) and achromatic conditions (gray diamonds). The error bars represent +/- 1 SEM taken between observers. The threshold 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 simulations. The parameters of the LINRF fit are provided in the legend. The data has been jittered for ease of viewing.

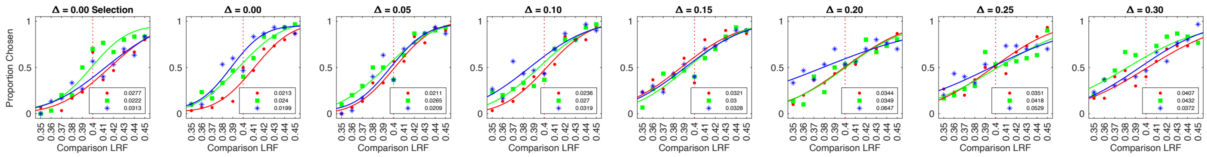

Figure 8: Psychometric functions for observer 0003 for light intensity variation experiment: Same as Figure 6, but for the light intensity variation experiment. The figure shows the proportion comparison chosen data for the selection session and the seven condition for observer 0003. The psychometric functions for all observers are shown in Figure S3.

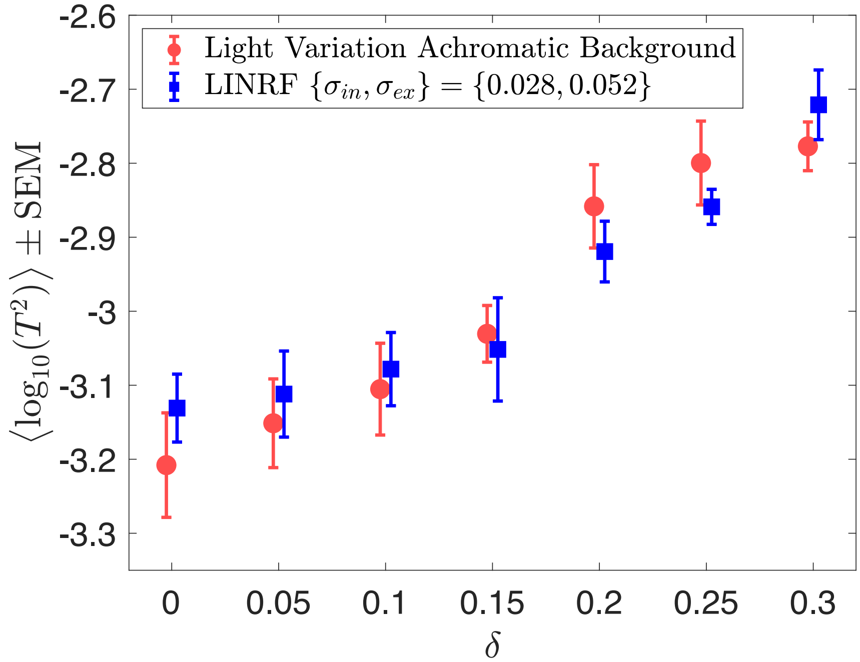

Figure 9: Light source intensity variation increases lightness discrimination threshold. Mean (N = 5) log squared threshold vs range parameter from human psychophysics for the seven light source intensity variation conditions (red circles). The error bars represent +/- 1 SEM taken between observers. 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 simulations. The parameters of the LINRF fit are provided in the legend. The data has been jittered for ease of viewing.

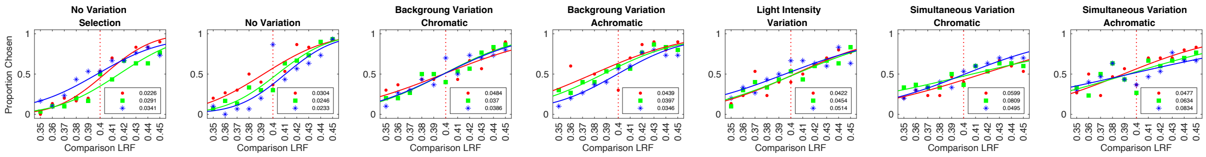

Figure 10: Psychometric functions for observer 0003 for simultaneous variation experiment: Same as Figure 6 and 8, but for simultaneous variation experiment. The figure shows the proportion comparison chosen data for the selection session and the six condition for observer 0003. The data for all observers are shown in Figure S5.

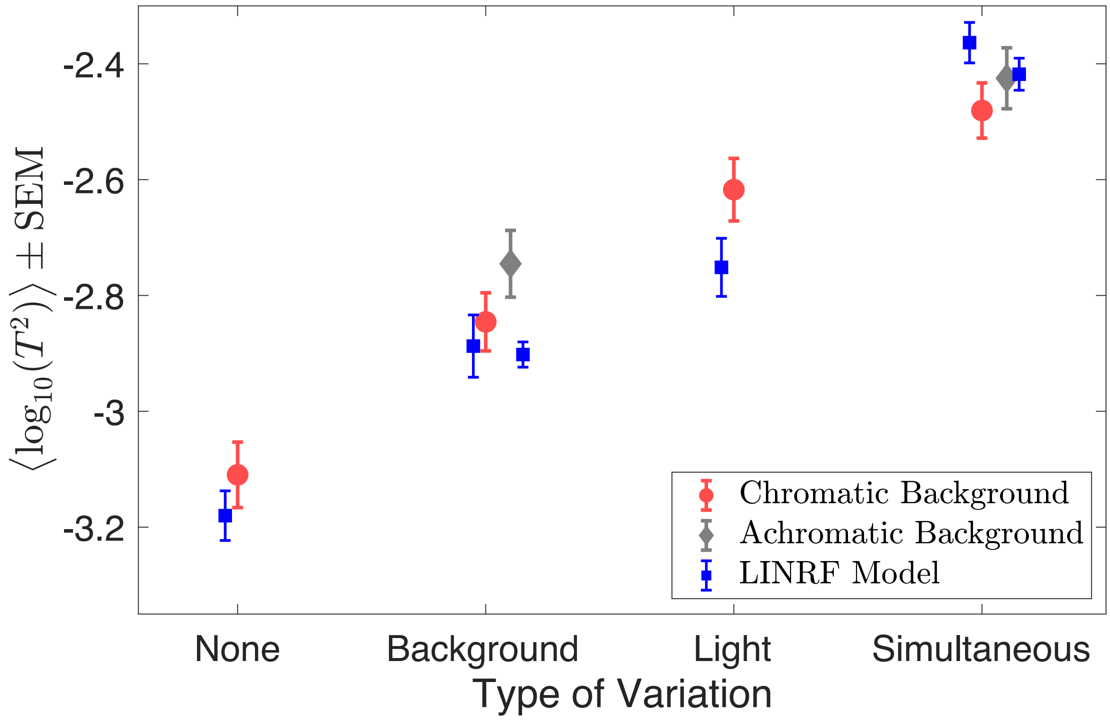

Figure 11: Discrimination thresholds for simultaneous variation of two sources are higher than individual discrimination thresholds. Mean (N = 6) log squared threshold for the six conditions in simultaneous variation experiment. The error bars represent +/- 1 SEM taken between observers. The data for 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 condition for (Figure 9) for the Light condition. The blue error bars show +/- 1 standard deviation estimated over 10 independent simulations.

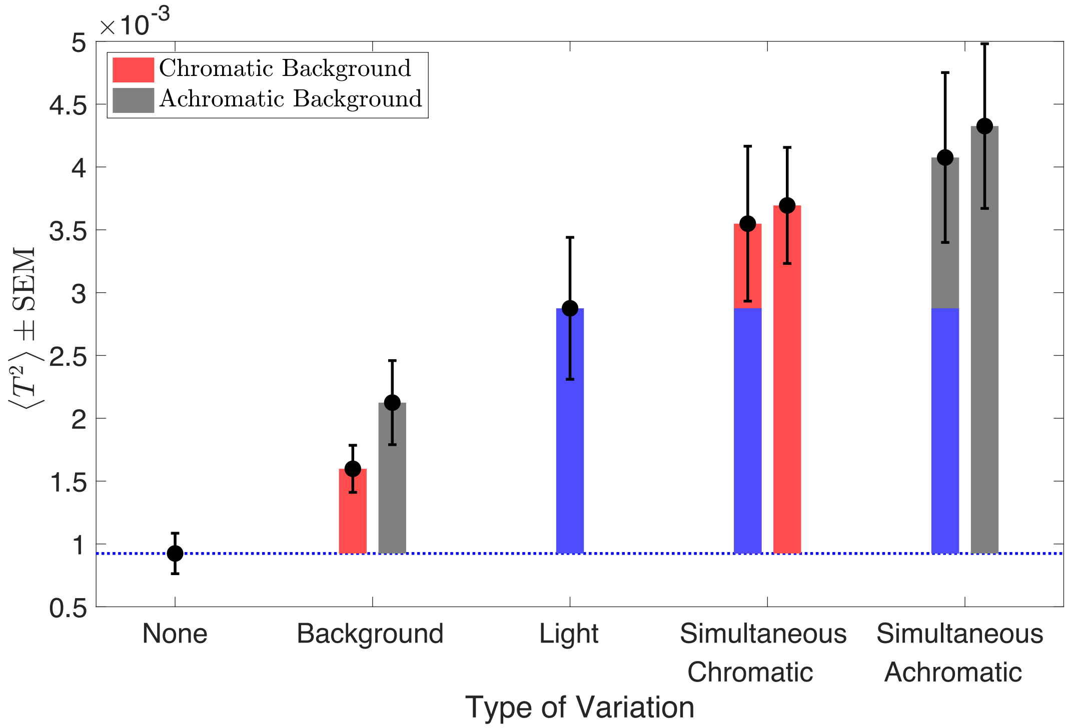

Figure 12: Extrinsic noise of independent variations add linearly for simultaneous variation: Mean squared thresholds (N=6) for the six conditions in simultaneous variation experiment (black circles). The black error bars represent +/- 1 SEM taken between observers. The bars (red, gray, blue) represent 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 measured squared threshold for the simultaneous variation conditions 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).

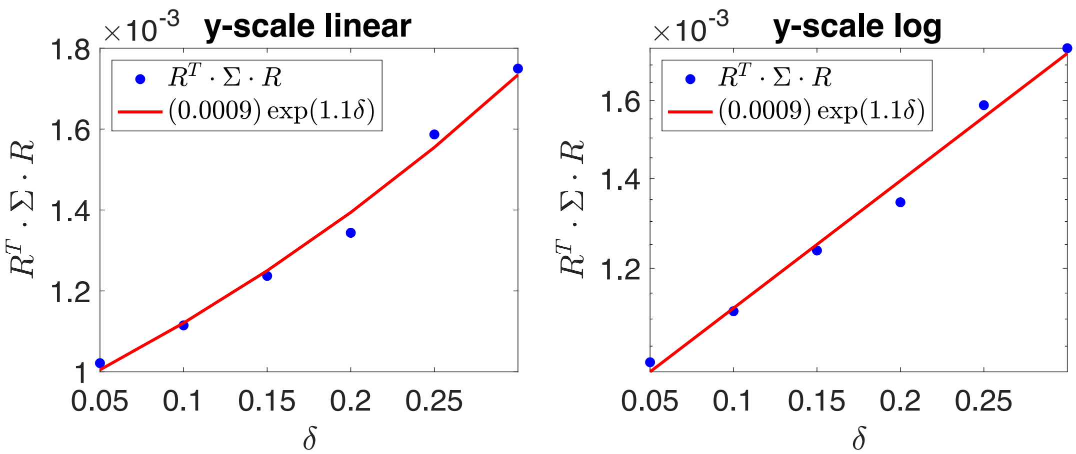

Figure S1: Estimation of extrinsic noise for light intensity variation: Plot of the variance ($R^T \Sigma R$) as a function of the range parameter δ on a linear (left panel) and logarithmic (right panel) scale. We fit the function with an exponential of the form $A * \exp{(B \cdot \delta)}$. The variance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

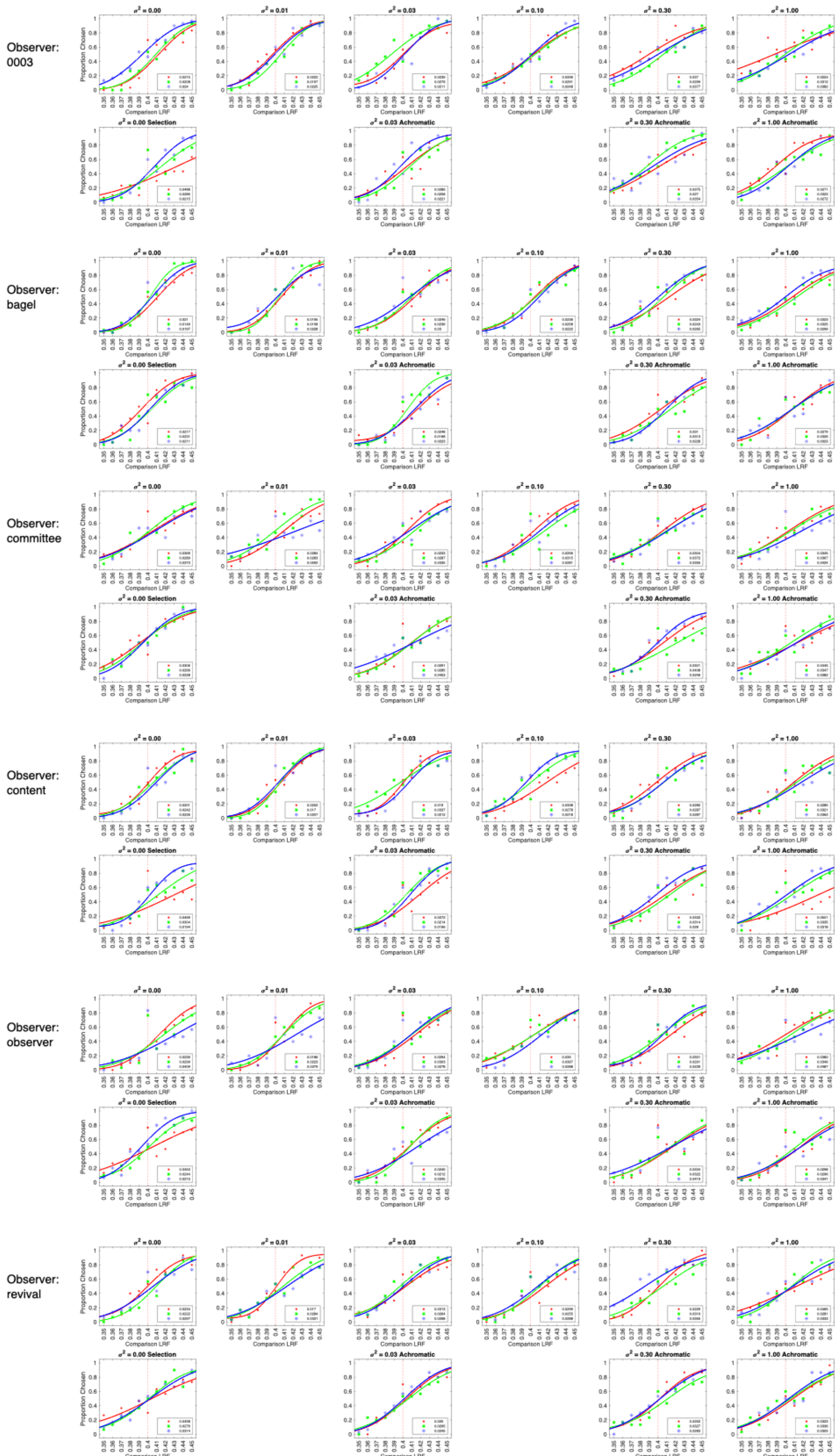

Figure S2: Psychometric functions for all observers for background variation experiment. Same as Figure 6, for all observers retained in background variation experiment.

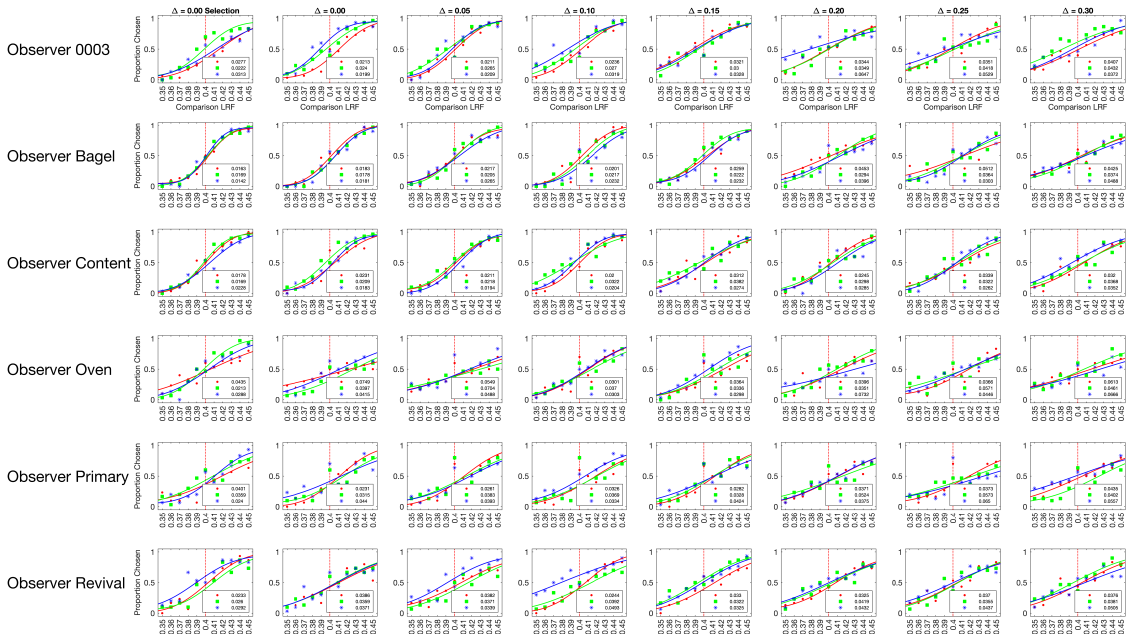

Figure S3: **Psychometric functions for all observers for light intensity variation experiment.** Same as Figure 8, for all observers retained in light intensity variation experiment.

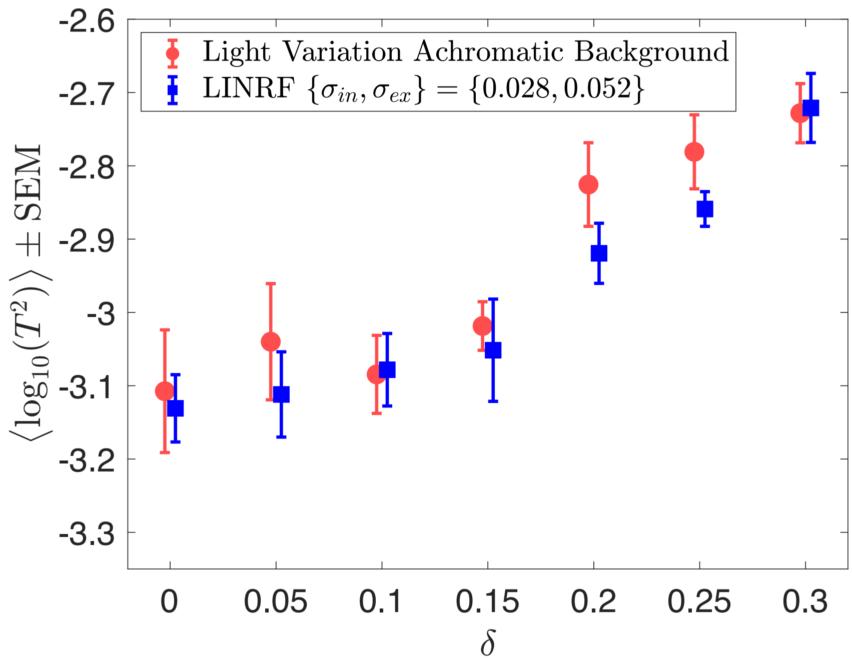

Figure S4: Same as Figure 9, for all six observers in retained in light intensity variation experiment. The parameters for the LINRF model are the same as in Figure 9.

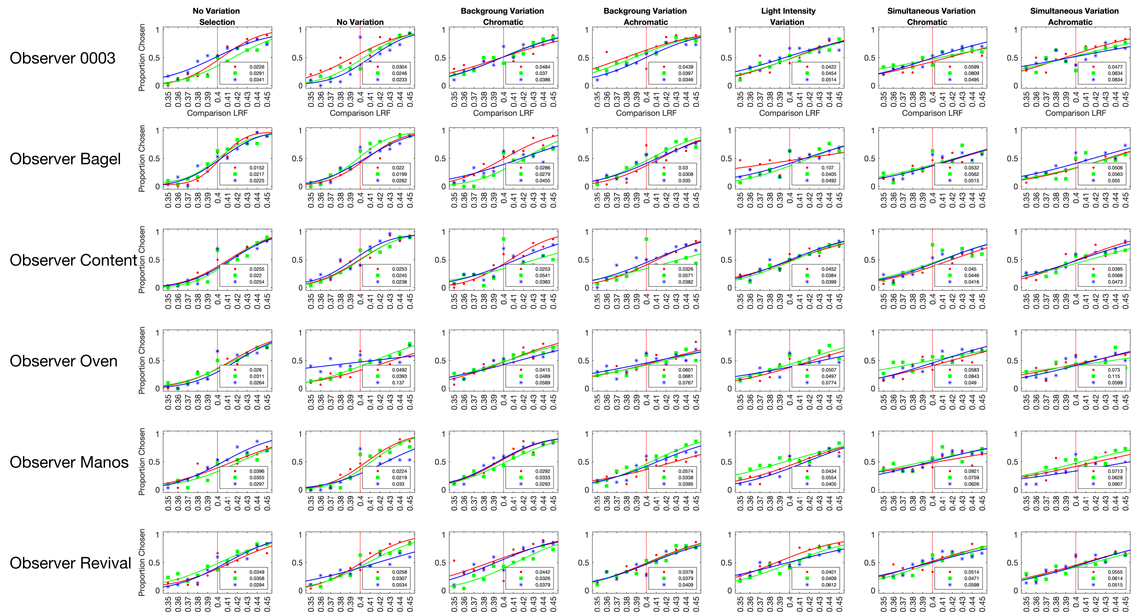

Figure S5: Psychometric functions for all observers for simultaneous variation experiment. Same as Figure 10, for all observers retained in simultaneous variation experiment.