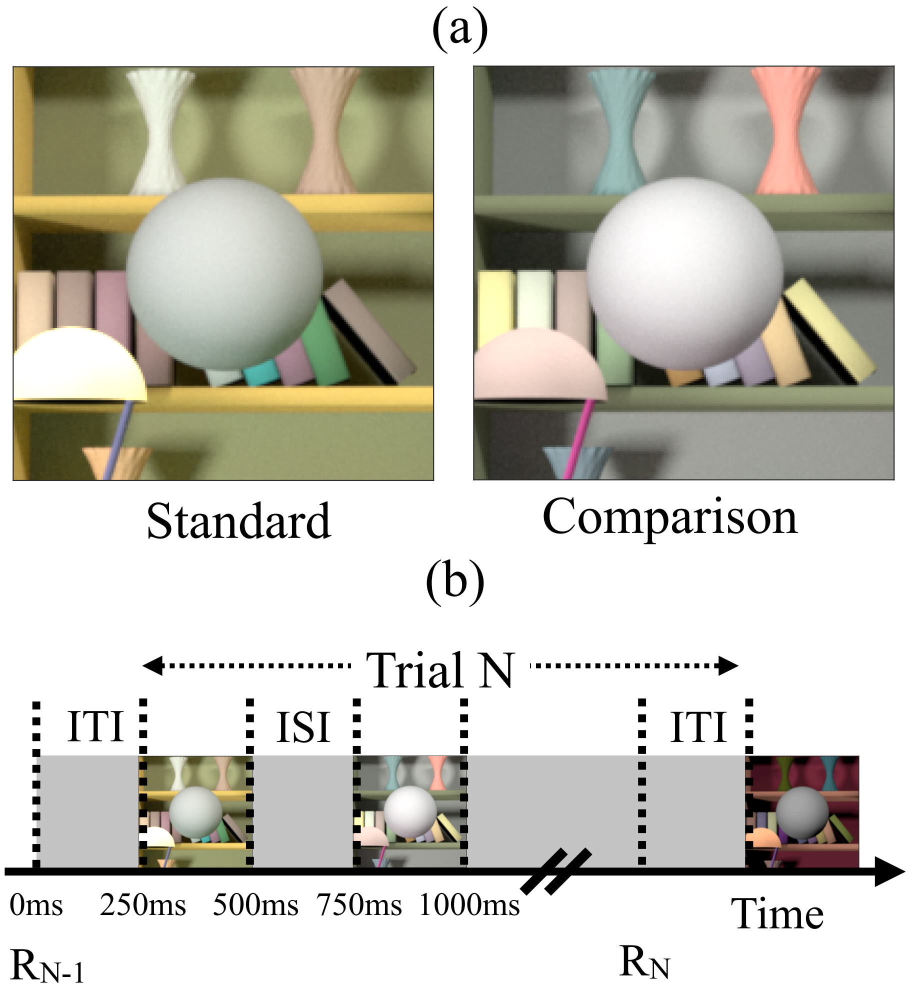

Figure 1: Psychophysical task. (a) On every trial of the experiment, human observers viewed two images in sequence, a standard image and a comparison image and indicated the one in which the spherical target object in the center of the image was lighter. Example standard and comparison images are shown. The images were computer graphics simulations. The simulated reflectance functions of the target were spectrally flat, and the spheres appeared gray. The overall reflectance of the target was held fixed in the standard images and differed between standard and comparison. Performance (proportion correct) was measured as a function of this difference to determine discrimination threshold. The reflectance functions of objects in the background could be held fixed or vary between standard and comparison on each trial (as illustrated here). The order of presentation of the standard and comparison images was randomized from trial to trial. Discrimination thresholds were measured as function of the amount of variation in background object reflectance.

(b) Trial sequence. R_{N-1} indicates the time of the observer's response for the (N-1)th trial. The Nth trial begins 250ms after that response (Inter Trial Interval, ITI). The Nth trial consists of two 250ms stimulus presentation intervals with a 250ms inter-stimulus interval (ISI). The observer responds by pressing a button on a gamepad after the second stimulus has been shown. The observer can take as long as he or she wishes before making the response, with an example response time denoted by R_N in the figure. The next trial begins 250ms after the response.

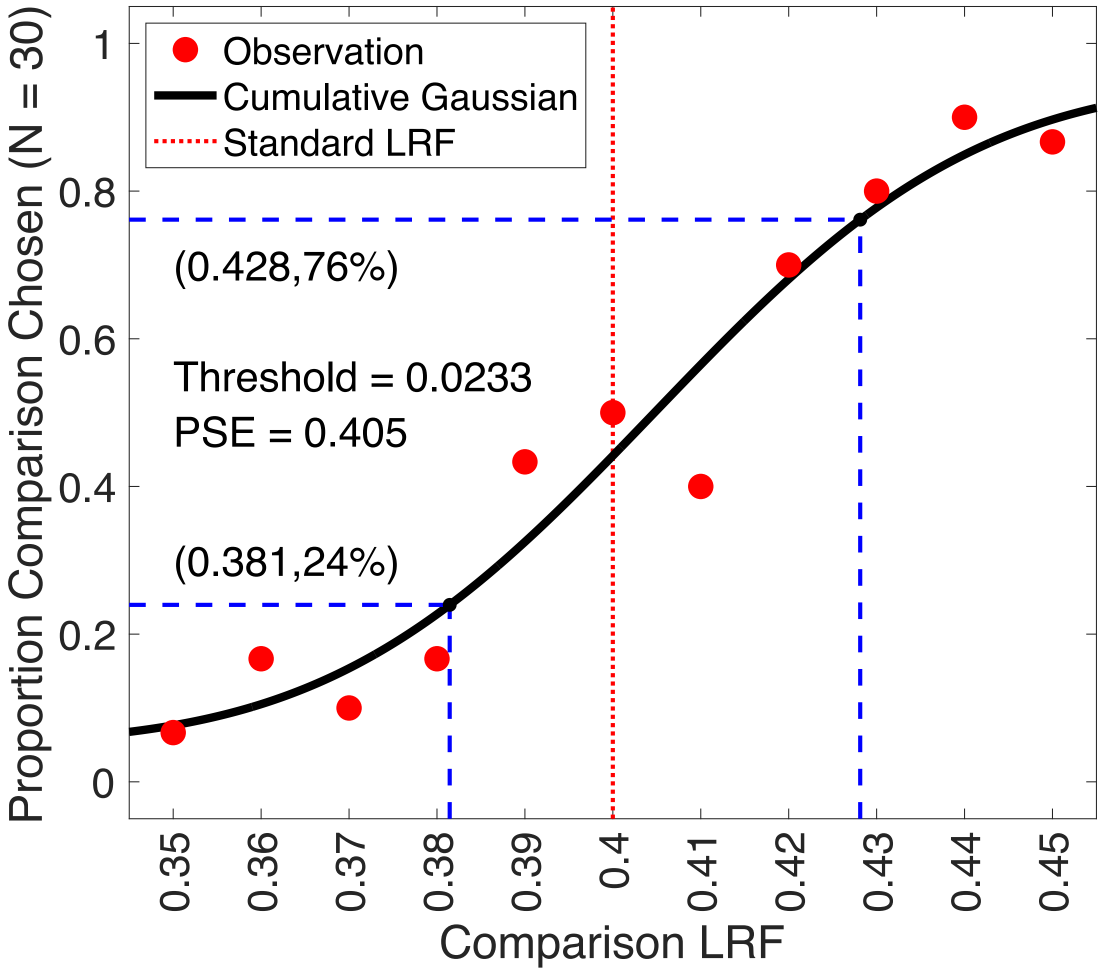

Figure 2: Psychometric function. We recorded the proportion of times the observer chose the target in the comparison image to be lighter, as a function of the comparison LRF. The LRF of the target object in the standard image was fixed at 0.4. The LRF of the target object in the comparison image were chosen from 11 linearly spaced values in the range [0.35, 0.45]. Thirty trials were presented at each comparison LRF value. We fit a cumulative normal distribution to the proportion comparison chosen data using maximum likelihood methods. The guess and lapse rates were assumed to be equal and were 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 0.7604 and 0.5, as predicted by the cumulative normal fit. This figure shows the data for Observer 2 for scale factor 0.00 in the first experimental session for that observer. The point of subjective equality (PSE, the LRF corresponding to proportion chosen 0.5) was close to 0.4 as expected and the threshold was 0.0233. The lapse rate for this fit was 0.05.

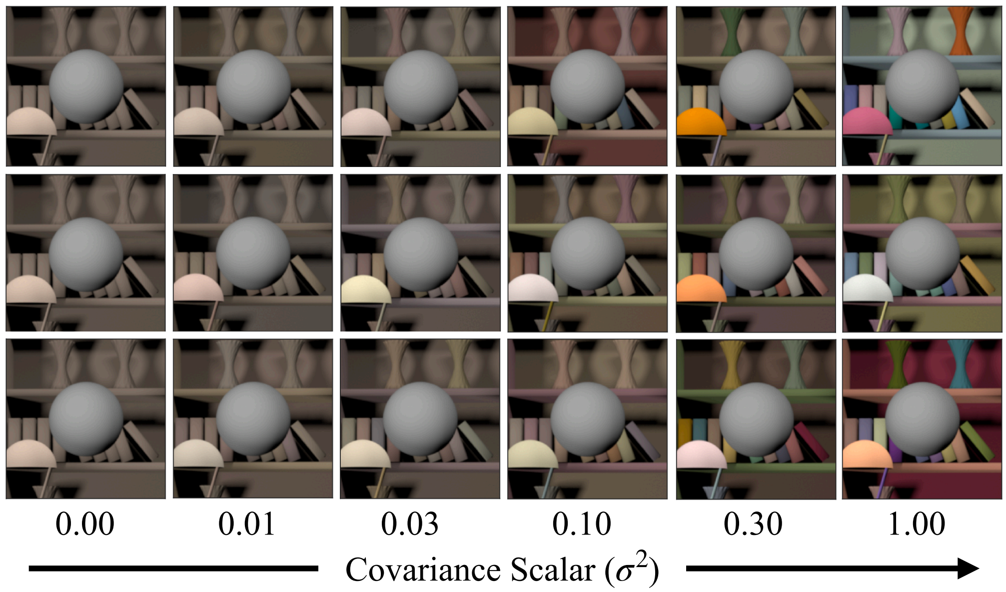

Figure 3: Variation in background surface reflectance: The reflectance spectra of background objects were chosen from a multivariate Gaussian distribution that modeled the statistics of natural surface spectra. The variation in the reflectance spectra was controlled by multiplying the covariance matrix of the distribution with a scalar. We generated images at six levels of the scalar. Each column shows three sample images at each of the six values of the scalar. The leftmost column corresponds to no variation and the rightmost column corresponds to the modeled variation of natural surfaces. The target object (sphere at the center of each panel) in each image has the same LRF. For each value of the scalar, we generated 1100 images, 100 each at 11 linearly spaced target LRF levels across the range [0.35, 0.45]. Discrimination thresholds were measured separately for each value of the covariance scalar shown.

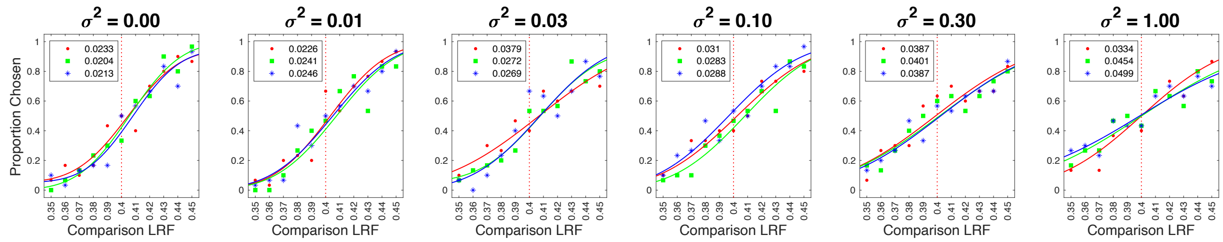

Figure 4: Change in psychometric functions with covariance scalar. Psychometric functions of an observer (Observer 2) for different amount of variation in the reflectance of background objects. We measured the proportion comparison chosen data at six values of the covariance scalar (σ^2), three times for each observer and at each value of covariance scalar. The data for each block was fit with a cumulative normal to obtain the discrimination threshold (See Figure 2). Each panel plots the measured values and the cumulative fit to the proportion comparison data in the three blocks. The values in the legend provide the estimate of lightness discrimination threshold for each block as obtained from the cumulative fit. See Figure S3 for the psychometric functions of all observers.

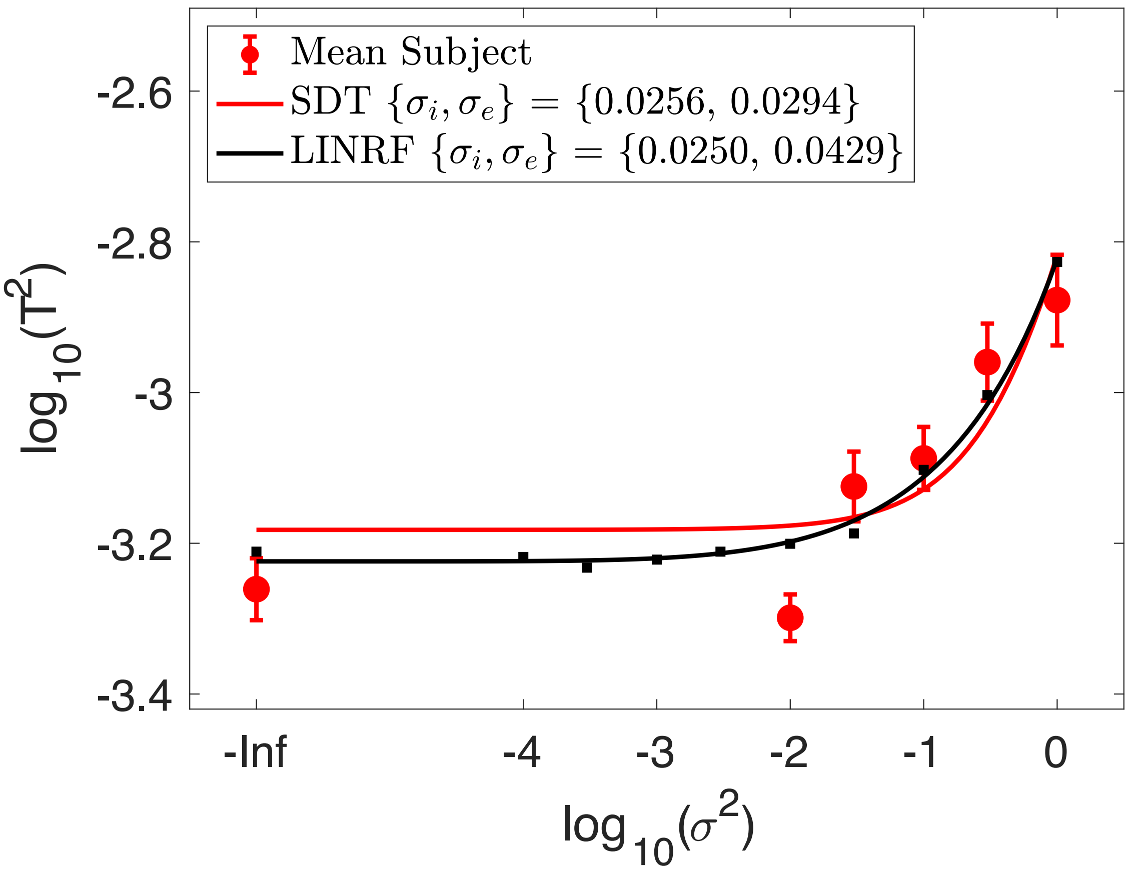

Figure 5: Background variation increases lightness discrimination threshold. Mean (N = 4) log squared threshold vs log covariance scalar from the human psychophysics (red circles). The error bars represent +/- 1 SEM taken between observers. The data were fit with the function (SDT Model) $T^2 = T_0^2\left(\sigma_i^2 + \sigma^2\sigma_e^2\right)$ with $T_0 = 1$ (red curve). The best fit parameters are indicated in the legend. The threshold of the linear receptive field (LINRF) model was estimated at 10 logarithmically spaced values of the covariance scalar (black squares). The black smooth curve is a smooth fit to these points of the functional form $\log_{10} T^2 = a + b^{(x+c)^d}$ where $x = \log_{10}\sigma^2$ and a, b, c and d are parameters adjusted in the fit.

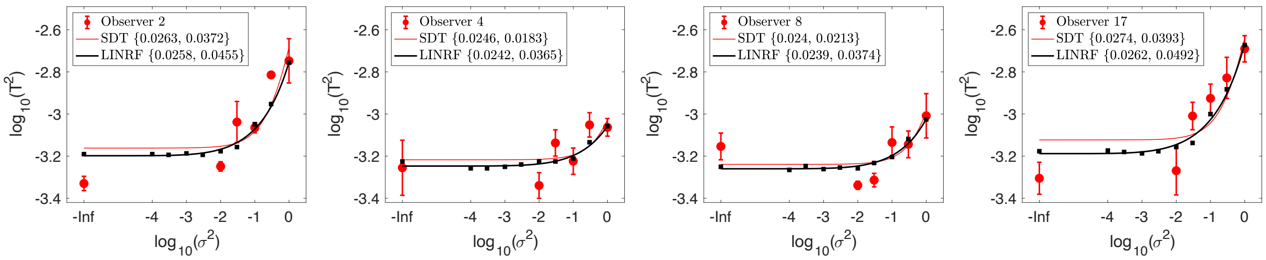

Figure 6: Threshold of individual human observers. Mean (across sessions) squared threshold vs log covariance scalar for individual human observers. Same format as Figure 4; here the error bars represent +/- 1 SEM taken across sessions for each observer. The parameters of the SDT model and the linear receptive field (LINRF) models were obtained separately for each observer.

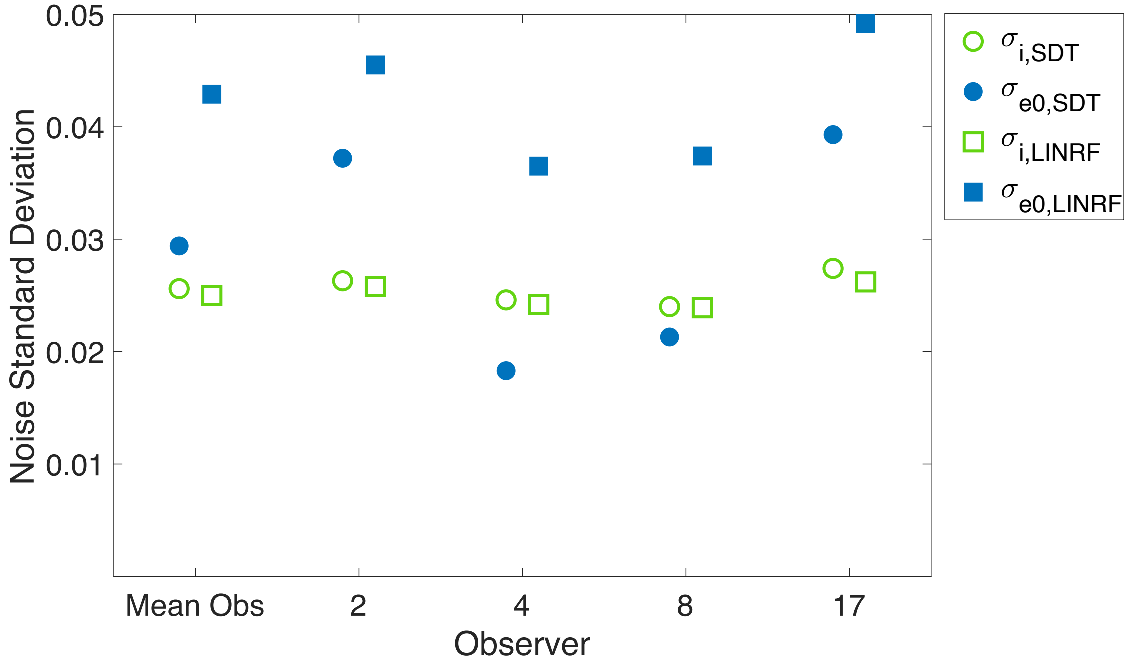

Figure 7: Internal and external noise standard deviation for human observers. Noise standard deviation for human observers estimated using SDT model and the computational linear receptive model (LINRF) model. While the internal noise estimates are consistent over the two models, the external noise estimated by the LINRF model is higher compared to the SDT model.

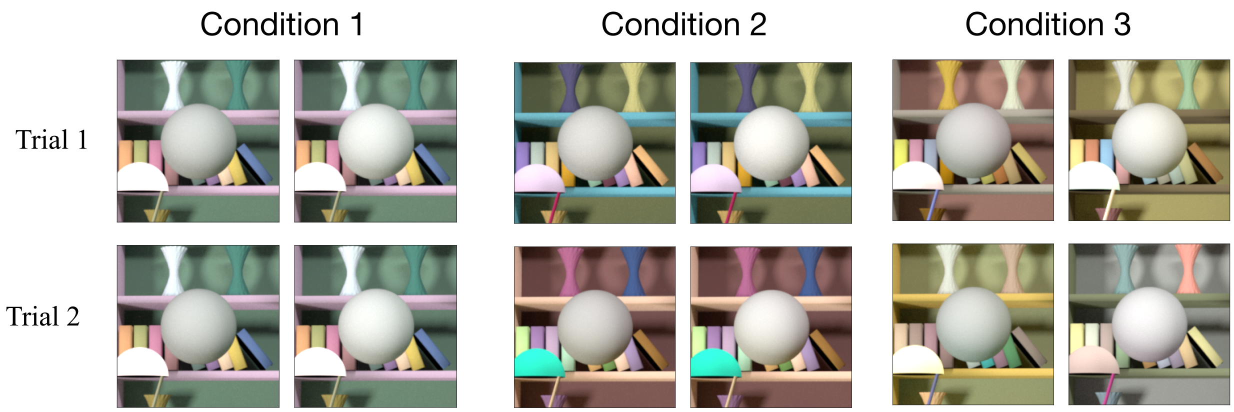

Figure S1: Example stimuli for Conditions 1, 2 and 3 in Experiment 2 to study the effect of background surface reflectance on lightness discrimination threshold. In condition 1, the background was fixed in every trail and every interval. In condition 2, the background varied from trial to trial, but remained fixed in the two intervals of a trial. In condition 3, the background varied in each trial and interval. For illustration, in this figure we have chosen the stimulus on the left to be the standard image with target object at 0.4 LRF and the on the right to be comparison image with target object at 0.45 LRF. In the experiment, the two images were presented sequentially in random order at the center of the screen. Conditions 2a and 3a stimuli are similar to condition 2 and 3 respectively, but without secondary reflections.

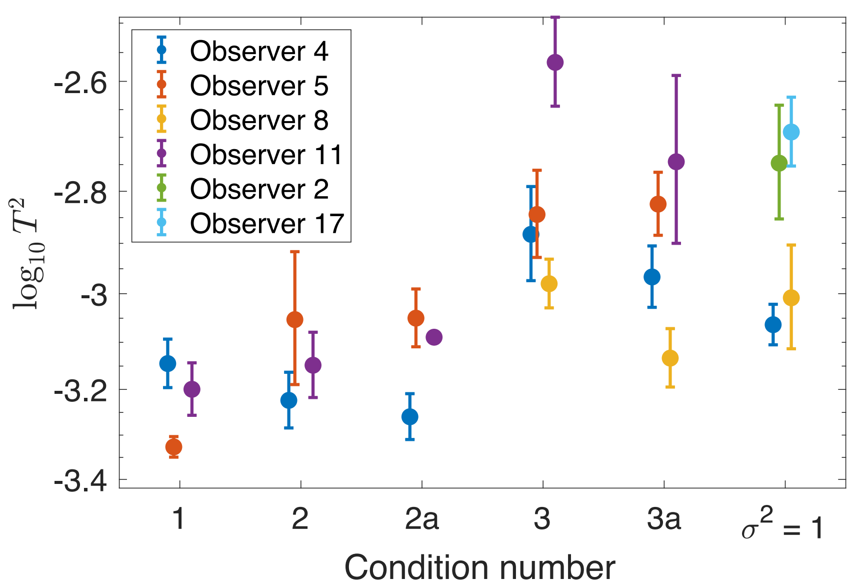

Figure S2: Lightness discrimination threshold of four human observers in the five conditions in Experiment 2 (The data points have been jittered to avoid marker overlaps). The thresholds are higher for the condition where the objects are compared against different backgrounds (Condition 3 and 3a) as compared to the same background (Condition 1, 2, 2a). Secondary reflections do not have any significant effect on thresholds (Condition 2a and 3a). Condition 3a of Experiment 2 is equivalent to the condition with covariance scalar equal to 1 ($\sigma^2 = 1$). The thresholds for this condition are also provided for comparison. Two observers from Experiment 2 also participated in Experiment 3.

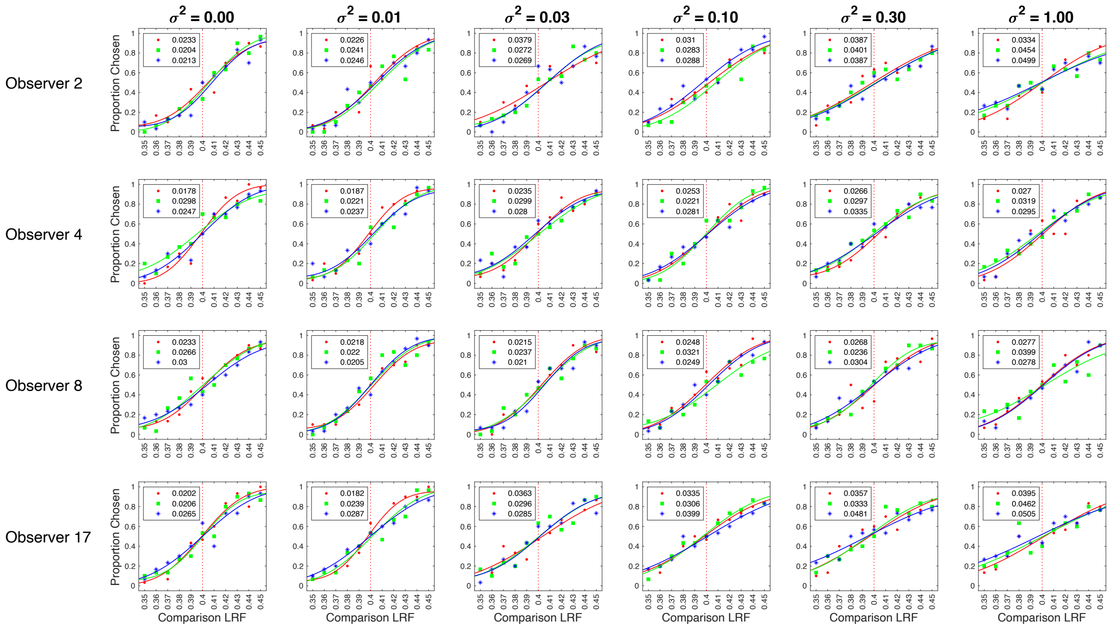

Figure S3: Change in psychometric functions with covariance scalar. Same as Figure 4 for all observers retained in the experiment.