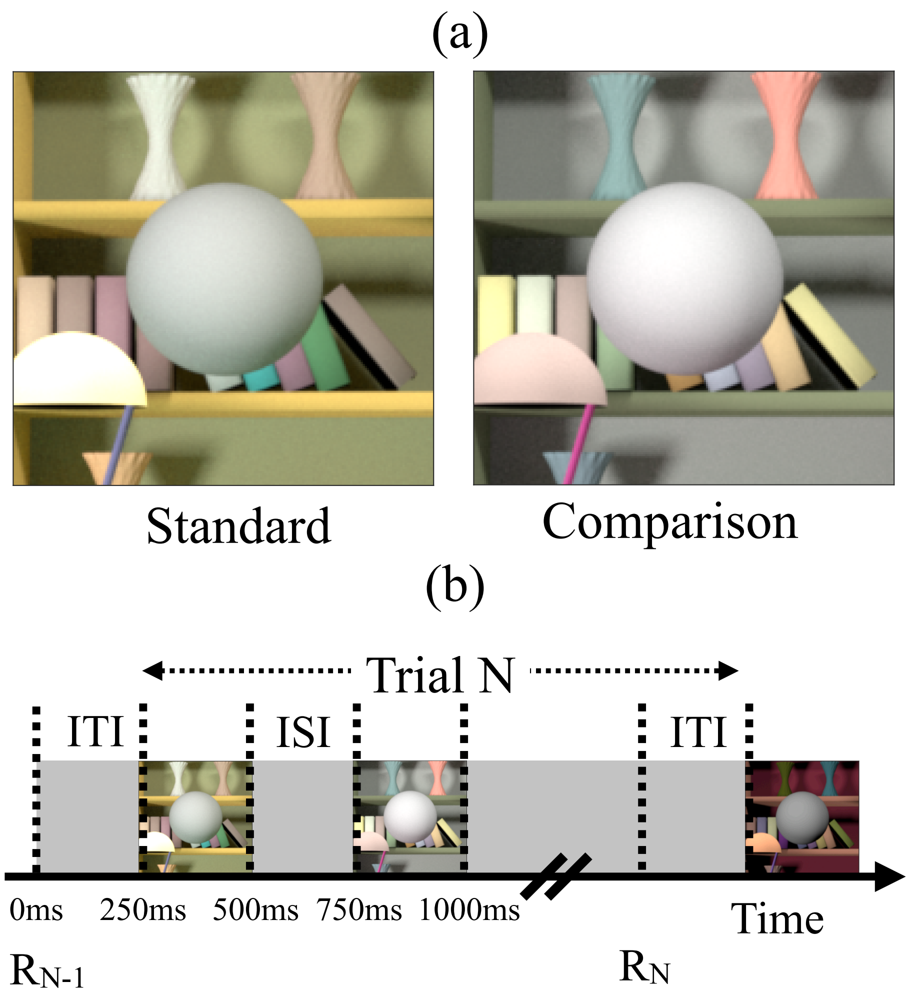

Figure 1: Psychophysical task. (a) On every trial of the experiment, human observers viewed two images in sequence, a standard image and a comparison image and indicated the one in which the spherical target object in the center of the image was lighter. Example standard and comparison images are shown. 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 N^{th} trial begins 250ms after that response (Inter Trial Interval, ITI). The N^{th} 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 time denoted by R_N in the figure. The next trial begins 250ms after the response.

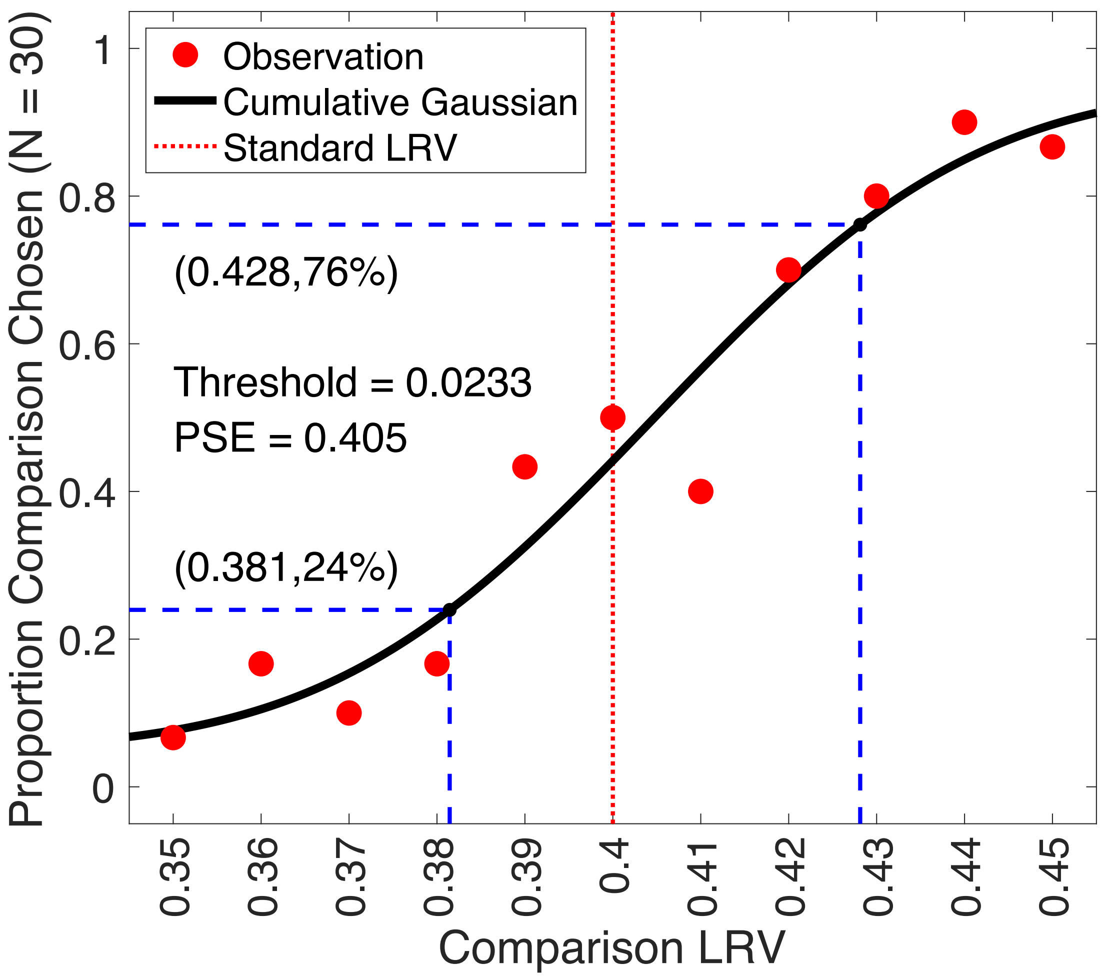

Figure 2: Psychometric function. We recorded the proportion of times the observer chose the target in the comparison image to be lighter, as a function of the comparison LRF. The LRF of the target object in the standard image was fixed at 0.4. The LRF of the target object in the comparison image were 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 CNSU_0002 for scale factor 0.00 in the first experimental session for that observer. The point of subjective equality (PSE, LRF for 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.

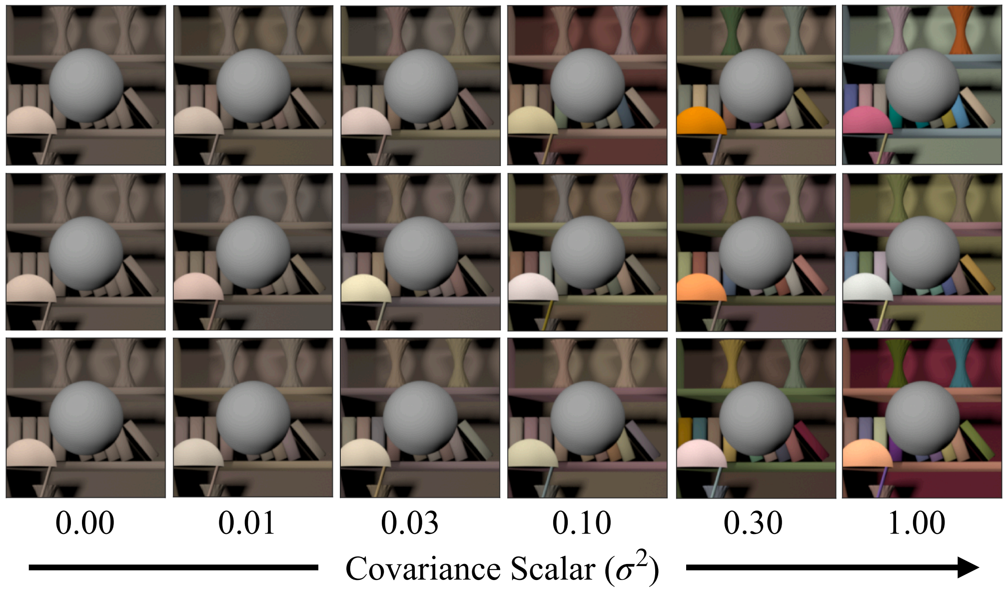

Figure 3: Variation in background color: The reflectance spectra of background objects were chosen from a statistical distribution of natural surfaces. The variation in the reflectance spectra was controlled by multiplying the covariance matrix of the distribution with a scalar. We generated images at 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 variation in natural surfaces. The target object (sphere at the center of each panel) in each image has the same lightness. For each value of the scalar, we generated 1100 images, 100 each at 11 linearly spaced lightness levels in the range [0.35, 0.45] LRF. Discrimination thresholds were measured separately for each value of the covariance scalar by presenting such images on a calibrated color monitor.

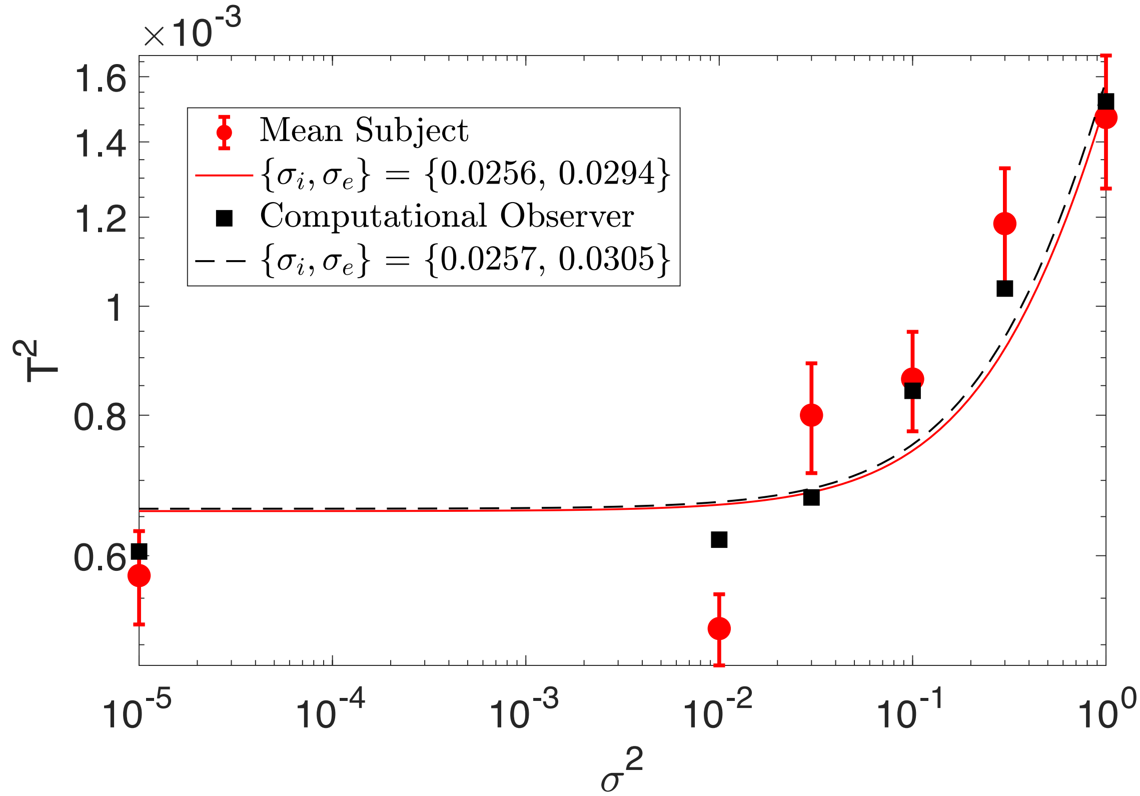

Figure 4: Background variation increases lightness discrimination threshold. Mean (N = 4) log squared threshold vs log covariance scalar from the human psychophysics (red circles). The error bars represent +/- 1 SEM taken between observers. The data were fit with the function $T^2 = T_0^2\left(\sigma_i^2 + \sigma^2\sigma_e^2\right)$ with $T_0 = 1$ (red curve). The best fit parameters are indicated in the legend. The predictions of the threshold of the computational model (black squares) and the corresponding model fit (black dashed line).

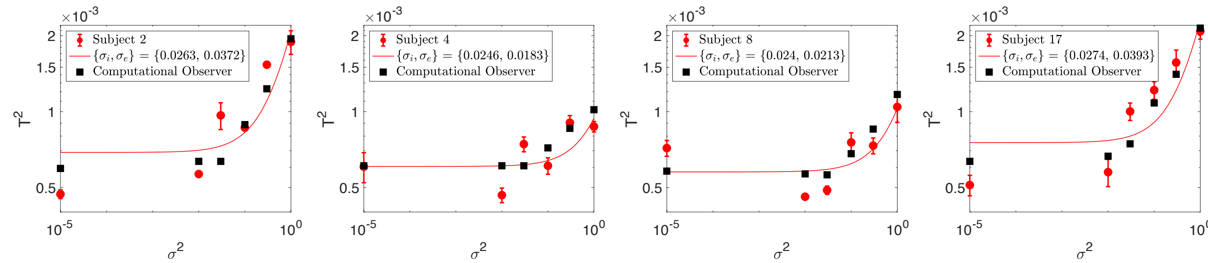

Figure 5: Threshold of individual human observers. Mean (across sessions) squared threshold vs log covariance scalar for individual human observers. Same format as Figure 4; here the error bars represent +/- 1 SEM taken across sessions for each observer. The smooth curves show the fit to the function $T^2 = T_0^2\ (\sigma_i^2 + \sigma^2\sigma_e^2)$, $(T_0 = 1)$. The parameters of the computational model were obtained separately for each observer by minimizing the mean square error between the mean observer thresholds and the predicted model thresholds for the six values of the covariance scalar.

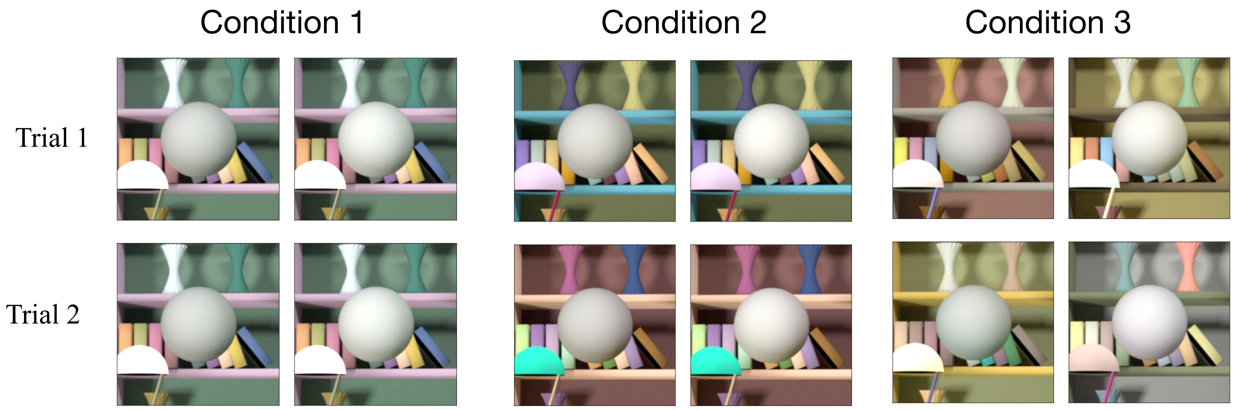

Figure S1: Two example trials of each condition in Experiment 2 to study the effect of background color on lightness discrimination threshold. In condition 1, the background was fixed in every trail and every interval. In condition 2, the background varied from trial to trial, but remained fixed in the 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.

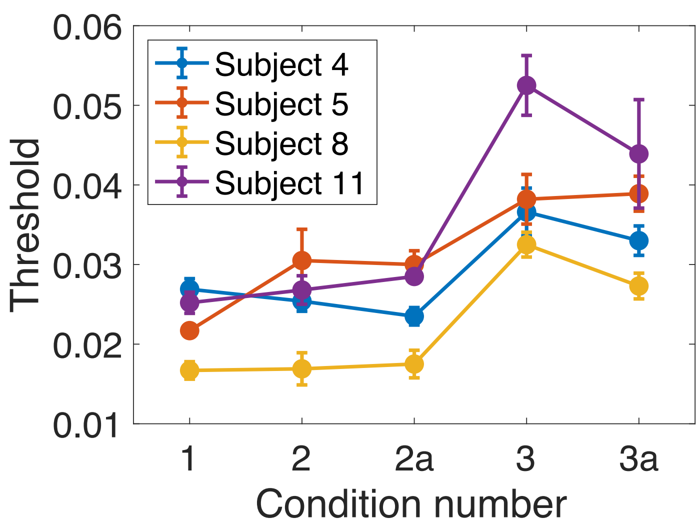

Figure S2: Lightness discrimination threshold of four observers in the Experiment 2. The discrimination thresholds are higher for the condition where the objects are compared against different backgrounds (Condition 3 and 3a) as compared to the same background (Condition 1, 2, 2a). Secondary reflections do not have any significant effect on thresholds (Condition 2a and 3a).