Title: lightness discrimination under spectral changes

Authors: Devin Reynolds, Vijay Singh.

Department of Physics, North Carolina Agricultural and Technical State University, Greensboro, NC, USA.

**ABSTRACT:**

**KEYWORDS:** Lightness, Human Psychophysics, Color Vision,

**INTRODUCTION**

**2 EXPERIMENTAL METHODS**

**Overview**

We followed the methodology as published in our previous work [cite paper]. In previous work, we measured human lightness discrimination thresholds under variability of reflectance spectra of background objects. The work presented here follows the same experimental methods, except that the stimuli used in the experiment were different. In this section, we will mainly focus on the differences from the previous work. We refer the reader to the previous work for details.

*We studied the effect of variability in object-extrinsic properties on the human ability to discriminate an object-intrinsic property. Specifically, we measured how variation in the reflectance spectra of background objects affects lightness discrimination thresholds, that is thresholds for discriminating object achromatic reflectance.[[1]](#footnote-1) We used a two-alternative forced-choice (2AFC) procedure (Figure 1). On each trial, observers viewed a standard image and comparison image, sequentially presented on a calibrated monitor for 250ms each. The inter-stimulus interval was 250ms (Figure 1a). The images were computer graphics renderings of 3D scenes. Each scene contained a spherical target object that appeared achromatic. The observers’ task was to report the image in which the target object was lighter. Across trials, we varied the luminous reflectance factor (LRF; American Society for Testing and Materials, 2017) of the target object in the comparison image while keeping the LRF of the target object in the standard image fixed. The LRF is the ratio of the luminance of a surface under a reference illuminant (here, the CIE D65 reference illuminant) to the luminance of the reference illuminant itself. The target object LRF was varied by scaling the surface reflectance spectrum of the target object, without changing its shape.[[2]](#footnote-2) The temporal order in which the standard and comparison images were presented was randomized on each trial.*

*We recorded the proportion of times observers chose the comparison image as having the lighter target object at 11 values of the target object LRF. Figure 2 shows a psychometric function from a typical human observer. The proportion-comparison-chosen data were fit with a cumulative normal using maximum likelihood methods (see Methods: Psychometric Function). Threshold was defined as the difference between the LRF of the target object at proportion comparison chosen 0.76 and 0.50 (i.e., d-prime = 1.0 in a two-interval task), as determined from the cumulative normal fit.*

*We measured lightness discrimination thresholds as a function of the amount of variability in the surface reflectances of the background objects in the rendered scenes. The reflectances of the background objects were chosen from a distribution of natural reflectances. The amount of variability was controlled parametrically by multiplying the covariance matrix of the distribution by a scalar (see Methods: Reflectance and Illumination Spectra). We measured thresholds for six logarithmically spaced values of this covariance scalar. By varying the scalar from 0 (no variation) to 1 (natural-scene-typical variation), we examined how background variation affects performance in the task. Figure 3 shows examples of images used in our psychophysical task for different choices of the covariance scalar.*

*The subsections below provide additional methodological detail.*

**Preregistration**

The experimental design and the method for extracting threshold from the data were preregistered before the start of the experiments. The preregistration documents are publicly available at: <https://osf.io/7tgy8/>.[[3]](#footnote-3)

We preregistered three experiments. The first experiment (preregistered as Experiment 6) studied the effect of variation in reflection spectra of background objects on human lightness discrimination thresholds. It was a replication of previous work (preregistered as Experiment 3, cite equivalent noise paper) with three additional conditions where the background objects were achromatic and varied only in their lightness. The second experiment (preregistered as Experiment 7) studied the variation in intensity of the light sources in the scene on human lightness discrimination thresholds. The third experiment (preregistered as Experiment 8) studied the variation of simultaneous variation of background object reflection spectra and the light source intensity on human lightness discrimination thresholds. The experimental methods of all three experiments were same.

We followed the procedure described in the preregistration document to extract threshold from the data. The document also indicated that the primary data feature of interest was the dependence of threshold on the amount of variation in the background and the intensity of the light source. We predicted that the thresholds would increase with increase in amount of variation. In the case of background variation, we predicted that the thresholds of achromatic variation would be lower than chromatic variation. We also predicted that increase in thresholds could be captured by our previously developed equivalent noise model. Additionally, we predicted that the threshold for simultaneous variation would be higher than the thresholds for individual variations.

**Reflectance and Illumination Spectra**

The reflectance spectra of background objects in the scene were generated using a model of naturally occurring surface reflectance spectra, as described in previous work. Briefly, we combine to datasets of surface reflectance functions containing 632 surface reflectance measurements. We then use principal component analysis (PCA) to obtain the projection of the mean centered dataset along the eigenvectors associated with the six largest eigenvalues. These eigenvalues captured more than 99.5% of the variance. We approximated the empirical distribution of the projection weights with a multivariate normal distribution. We generated pseudorandom samples from this multivariate normal distribution to get the projection weights of random samples of reflectance spectra. Reflectance spectra were constructed by using these projection weight along with the eigenvectors and adding the mean of the surface reflectance dataset. A physical realizability condition was imposed on these spectra by ensuring that the reflectance at each wavelength was between 0 and 1. If a reflectance spectrum did not meet this criteria, it was discarded.

To generate achromatic surface reflectance spectra, after generating a physically realizable reflectance spectrum, its average reflectance over all wavelengths was calculated and it was replaced by a spectrum which had this average reflectance all wavelengths.

To control the amount of variation in the reflectance spectra, the covariance matrix of the multivariate normal distribution was multiplied by a covariance scalar. A covariance scalar of 0 corresponds to no background object reflectance variation. A covariance scalar of 1 corresponds to the full reflectance variation of the model of natural reflectance.

The power spectrum of the light sources was chosen as that of standard daylight D65. We normalized the D65 spectrum by its mean power to obtain its relative spectral shape. The variation in the light source intensity was introduced by multiplying the normalized D65 spectrum by a random sample generated from a log-uniform distribution in the range [1−Δ, 1+Δ], where the parameter Δ determines the range of the distribution. We chose log-uniform distribution for the multiplication parameter because the spectral power distribution function of natural daylight spectra varies over three orders of magnitude and their mean over wavelength can be roughly approximated by a log-uniform distribution (cite VWCC paper). All light sources in a scene were assigned the same power spectrum.

Background reflectance variation (preregistered Experiment 6): To study lightness discrimination thresholds with variation in reflectance spectra of background objects, we generated images for nine conditions. Six of these conditions were for chromatic variation at six logarithmically spaced values of covariance scalar: [0, 0.01, 0.03, 0.1, 0.3, 1.0]. Three conditions were for achromatic variation at covariance scalars: 0.03, 0.3 and 1.0. The power spectrum of the light source was the same for all images. The multiplication scalar was assigned an arbitrary value of 5.

When displayed on the experimental monitor, the average luminance of the standard image for covariance scalar 0.00 was 94.0 cd/m2. The average luminances of the target object for the 11 LRF levels were [130.0, 131.5, 133.0, 134.5, 136.0, 137.5, 138.9, 140.3, 141.7, 143.1, 144.5] cd/m2.

Light source intensity variation (preregistered Experiment 7): To study lightness discrimination thresholds with variation in light source intensity, we generated images for seven values of the range parameter: [0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30]. The reflectance spectra of all background objects were the same and was equal to the mean spectrum of the reflectance database. This corresponds to covariance scalar of 0.

When displayed on the experimental monitor, the average luminance of the standard image for covariance scalar 0.00 and range parameter 0.00 was 93.3 cd/m2. The average luminances of the target object for the 11 LRF levels were [128.9, 130.5, 132.0, 133.5, 135.0, 136.4, 137.9, 139.3, 140.7, 142.1, 143.4] cd/m2.

Simultaneous variation (preregistered Experiment 8): In this experiment we studied seven conditions. These were: no variation (covariance scalar = 0, range parameter = 0), chromatic background variation (covariance scalar = 1, range parameter = 0), achromatic background variation (covariance scalar = 1, range parameter = 0),light source intensity variation (covariance scalar = 0, range parameter = 0.3), and simultaneous variation chromatic background (covariance scalar = 1, range parameter = 0.3) and simultaneous variation achromatic background (covariance scalar = 1, range parameter = 0.3).

When displayed on the experimental monitor, the average luminance of the standard image for covariance scalar 0.0 and range parameter 0.00 was 87.1 cd/m2. The average luminances of the target object for the 11 LRF levels were [120.9, 122.3, 123.8, 125.2, 126.5, 127.9, 129.2, 130.5, 131.9, 133.1, 134.4] cd/m2.

When displayed on the experimental monitor, the average luminance of the standard image for covariance scalar 1.0 and range parameter 0.30 was 87.8 cd/m2. The average luminances of the target object for the 11 LRF levels were [117.7, 119.4, 119.4, 122.3, 123.7, 123.8, 127.8, 126.9, 127.7, 129.1, 129.0] cd/m2.

***Image Generation***

*The images were generated using software we refer to as Virtual World Color Constancy (VWCC) (*[*github.com/BrainardLab/VirtualWorldColorConstancy*](https://github.com/BrainardLab/VirtualWorldColorConstancy)*). VWCC is written using MATLAB. It harnesses the Mitsuba renderer (Jakob, 2010) to render simulated images from scene descriptions, and also takes advantage of our RenderToolbox package (rendertoolbox.org; Heasly, Cottaris, Lichtman, Xiao, & Brainard, 2014). To render an image, we first create a 3D model that specifies the base scene. Objects and light sources can be inserted in the base scene at user specified locations. The 3D models utilized a base scene provided as part of RenderToolbox and modified using Blender, an open-source 3-D modeling and animation package (blender.org). Next, we assigned reflectance spectra and spectral power distribution functions to the objects and light sources in the scene (see Methods: Reflectance and Illumination Spectra). For each image, reflectances were assigned to the background objects by random draw from the reflectance model described above, with appropriate covariance scale factor. This procedure means that a set of images embodies the variation in background spectra described by the reflectance model, with each individual image containing a variety of background reflectances (Figure 3). Illumination spectra were not varied throughout the experiments reported here, and illumination spectra were as described in Methods: Reflectance and Illumination Spectra above.*

*Once the geometrical and spectral features were specified, we rendered a 2D multispectral image of the scene using Mitsuba, a physically-realistic open-source rendering system (mitsuba-renderer.org; Jakob, 2010). The images were rendered at 31 wavelengths equally spaced between 400nm and 700nm. The images were rendered with the camera field of view of 17° with an image resolution of 320-pixel by 240-pixels with the target object at the center. A 201-pixel by 201-pixel area, centered around the spherical target object, was cropped for display on the monitor.*

*To present the multispectral images on the monitor, they were first converted to LMS images using the Stockman-Sharpe 2° cone fundamentals (T\_cones\_ss2 in the Psychophysics Toolbox). Then the monitor calibration data and standard methods (Brainard, 1989; Brainard, Pelli, & Robson, 2002) were used to convert the LMS images to gamma corrected RGB images. A common scaling was applied to all images before rendering to ensure that they were within monitor gamut, so that the maximum linear channel RGB channel input was 0.9. The gamma corrected RGB images was presented on the monitor during the experiment.*

***Stimulus Design***

*As noted above, we measured lightness discrimination thresholds for six values of the covariance scalar. For each value of the covariance scalar, we generated a dataset of 1100 images. The dataset had 100 images each at 11 values of the target object LRF. The LRF of the target object in the standard images was 0.4 and the LRF in the comparison image varied between 0.35 and 0.45 at steps of 0.01 (11 comparison levels). We generated 100 images at each comparison level, each with a different choice of the reflectance spectra of the background objects. The fact that we had 100 images for each target LRF allowed us to randomize the background object reflectances across the two intervals of each forced choice trial without excessive replication. For covariance scalar 0.00 we generated a set of 11 images, one at each LRF level, as the background remained fixed in this case. All images were generated without secondary reflections specified in the rendering process. The geometry of the 3D scene was also held fixed across all images.*

**Experimental Details**

*We define a trial as the presentation of two images (standard and comparison images) and collection of the observer’s response. We define an interval as the presentation of one of the images in the trial.*

*The experiment was structured as follows. We define a block of trials as the data collected at one covariance scalar with 30 trials at each of the 11 comparison levels. We define a permutation as a set of six blocks, where each block corresponds to one of the possible six covariance scalars. We collected three permutations for each observer, with a new random order drawn for each permutation. Thus, after the practice session (see Methods: Observer Recruitment and Exclusion), there were total 18 blocks. We divided these 18 blocks over 6 sessions, each session with 3 blocks. In each block, we randomly selected the images for the trials from the pre-generated image database. The first five trials of each block were moderate trials (as defined in Methods: Observer Recruitment and Exclusion) to acclimatize the observer to the experimental task. The responses for these five trials were not saved.*

*The trial sequence (comparison level, specific images, standard/comparison order) in a block was generated pseudo-randomly at the beginning of the block. For this, at each comparison lightness level, 30 standard and comparison images were chosen pseudo-randomly with replacement from the image dataset. The sequence of presentation of these 330 trials were randomized and saved. For each trial, the order of presentation of the standard and comparison image was also determined pseudo-randomly and saved. The trials were presented according to the saved sequence.*

*The trials in a block were presented in three sub-blocks of 110 trials each. At the end of each sub-block the observer took a break of minimum duration 1 minute. The observer could terminate the experiment anytime during the block. If an observer terminated a block, the data for that block was not saved. No observer terminated any block. One observer indicated a desire to postpone at the beginning of a session, due to fatigue for reasons unrelated to the experiment. The session was rescheduled.*

*At the beginning of the first experimental session (the practice session) for each observer, the experimenter explained the experimental procedures and obtained consent for the experiments. The experimenter then tested the observer for normal visual acuity and color vision. The observer was then taken to the experimental room, where the experimenter described the task, and the observer was shown the display, chin rest, and response box. The observer was dark adapted by sitting in the dark room for approximately 5 minutes. The observer then performed the familiarization block (see Methods: Observer Recruitment and Exclusion for explanation of familiarization block). After the familiarization block, the observer performed the other three blocks of the practice session. The practice session lasted about one hour.*

*Observers who met the inclusion criteria (see Methods: Observer Recruitment and Exclusion) then performed 18 blocks over 6 additional sessions, each on a separate day. The order of blocks for each observer was determined pseudo-randomly at the beginning of the practice session. As noted above, observers performed three blocks per session. Observers were dark adapted for 5 minutes at the beginning of each session. The data for all observers in the main experiment (preregistered Experiment 3) were collected over a period of four weeks.*

*Observers viewed the stimuli with both eyes.*

**Observer Recruitment and Exclusion**

Observers were recruited from the North Carolina Agricultural and Technical State University and the local Greensboro community and were compensated for their time. Observers were screened to have normal visual acuity (20/40 or better; with corrective eyewear, if applicable) and normal color vision, as assessed with pseudo-isochromatic plates (Ishihara, 1977). These exclusion criteria were specified in the preregistration document (see Methods: Preregistration).

Observers who passed the vision screening then participated in a practice session. This session also served to screen for observers’ ability to reliably perform the psychophysical task. The observer was excluded from further participation if their mean threshold for the last two blocks in the practice session exceeded 0.030 (log T2, -3.2). This exclusion criterion was specified in our preregistered protocol (See Methods: Preregistration).

Observers who met the performance criterion participated in the rest of the experiment.

**Observer Information**

Background reflectance variation (preregistered Experiment 6): A total of 25 observers participated in the practice sessions for background variation experiment (10 Female, 15 Male; age 19-34; mean age 22.9). To de-identify observer information in the data, observers were given pseudo-names chosen by the experimenter. Six of these observers (pseudo-names: *0003, bagel, committee, content, observer*, and *revival*) met the performance criterion set for screening (2 Female, 4 Male; age 19-28; mean age 23.33). All observers who advanced to the practice session had normal or corrected-to-normal vision (20/40 or better in both eyes, assessed using Snellen chart) and normal color vision (0 Ishihara plates read incorrectly). The visual acuities of the observers in the main experiment were: *0003*, L = 20/30, R = 20/20; *bagel*, L = 20/20, R = 20/20; *committee*, L = 20/25, R = 20/25; *content*, L = 20/20, R = 20/20; *observer*, L = 20/25, R = 20/25; *revival*, L = 20/20, R = 20/20. *Committee, content,* and *observer* wore personal corrective eyewear both during vision testing and during the experiments. Observers *0003, bagel*, and *revival* did not require or use corrective eyewear.

Light source intensity variation (preregistered Experiment 7): A total of 15 observers participated in the practice sessions for light source intensity variation experiment (9 Female, 6 Male; age 19-33; mean age 25). Six of these observers (pseudo-names: 0003, bagel, content, oven, primary, and revival) met the performance criterion set for screening (3 Female, 3 Male; age 19-28; mean age 23.83). All observers who advanced to the practice session had normal or corrected-to-normal vision (20/40 or better in both eyes, assessed using Snellen chart) and normal color vision (0 Ishihara plates read incorrectly). The visual acuities of the observers in the main experiment were: *0003*, L = 20/30, R = 20/30; *bagel*, L = 20/20, R = 20/20; *content*, L = 20/20, R = 20/20; *oven*, L = 20/20, R = 20/20; *primary*, L = 20/20, R = 20/20; *revival*, L = 20/20, R = 20/20. Observer *Content* and *primary* wore personal corrective eyewear both during vision testing and during the experiments. Observers *0003, bagel*, *oven,* and *revival* did not require or use corrective eyewear. Observer *oven* reported some difficulties during a few sessions of the experiment and their thresholds for two conditions did not fit the expected pattern. We removed their data from the analysis presented in this work. Their data is provided in supplementary materials.

Simultaneous variation (preregistered Experiment 8): A total of 20 observers participated in the practice sessions for simultaneous variation experiment (9 Female, 11 Male; age 19-28; mean age 20.8). Six of these observers (pseudo-names: *0003, bagel, oven, fun, manos,* and *revival*) met the performance criterion set for screening (2 Female, 4 Male; age 19-28; mean age 23.33). All observers who advanced to the practice session had normal or corrected-to-normal vision (20/40 or better in both eyes, assessed using Snellen chart) and normal color vision (0 Ishihara plates read incorrectly). The visual acuities of the observers in the main experiment were: *0003*, L = 20/30, R = 20/30; *bagel*, L = 20/20, R = 20/20; *oven*, L = 20/20, R = 20/20; *fun*, L = 20/20, R = 20/20; *manos*, L = 20/25, R = 20/25; *revival*, L = 20/20, R = 20/20. *Oven*, and *fun* wore personal corrective eyewear both during vision testing and during the experiments. Observers *0003, bagel*, *manos*, and *revival* did not require or use corrective eyewear.

**Apparatus**

The stimuli were presented on a calibrated LCD color monitor (27-in. NEC MultiSync EA271U; NEC Display Solutions) in an otherwise dark room. The monitor was driven at a pixel resolution of 1920 x 1080, a refresh rate of 60Hz, and with 8-bit resolution for each RGB channel. The host computer was an Apple Macintosh with an Intel Core i7 processor. The experimental programs were written in MATLAB (MathWorks; Natick, MA) and relied on routines from the Psychophysics Toolbox (<http://psychtoolbox.org>) and mgl (<http://justingardner.net/doku.php/mgl/overview>). Responses were collected using a Logitech F310 gamepad controller.

The observer’s head position was stabilized using a chin cup and forehead rest (Headspot, UHCOTech, Houston, TX). The observer’s eyes were centered horizontally and vertically with respect to the display. The distance from observer’s eyes to the monitor was 75cm.

**Monitor Calibration**

The monitor was calibrated using a spectroradiometer (PhotoResearch PR655) as described in [cite previous paper.]. The monitor was calibrated before starting each experiment. Once calibrated the same settings were used till data for all observers for that experiment was collected. The monitor was then recalibrated for the next experiment. Data was collected in the sequence preregistered Experiment 6, Experiment 7, and Experiment 8.

*The monitor was calibrated using a spectroradiometer (PhotoResearch PR650). To calibrate the monitor, we focused the spectroradiometer on a patch displayed on the center of the monitor. The patch size was 4.66cm x 4.66cm (3.56° x 3.56°). The optics of the radiometer sampled the emitted light from a 1° circular spot within the patch. The spectral power distribution of the three monitor primaries was measured in the range 380nm to 780nm at 4nm steps. The gamma functions for each primary were determined from measurements of the spectral power distribution for each primary at 26 equally spaced input values for that primary, in the range [0, 1] where 1 corresponds to the maximum input value of the device. These gamma functions as well as the light emitted by the monitor for an input of 0 were accounted for in the stimulus display procedures. The spectral power distribution was also measured for 32 different combinations of RGB input values. These measurements were used to check the performance of the display. The maximum absolute deviation of the x-y chromaticity between the measured values and those predicted from the calibration was 0.0028 and 0.0027 for x and y chromaticity respectively, and less than 1% for luminance.*

**Stimulus Presentation**

*The size of each image was 2.6cm x 2.6cm on the monitor, corresponding to 2° by 2° visual angle. The target object size on the screen in the 2D images was ~1° in diameter. Each image was presented for 250ms (this was a deviation from the preregistration document, which specifies the presentation time as 500ms), with an inter-stimulus interval of 250ms and inter-trial interval of 250ms. Inter-stimulus interval (ISI) is defined as the interval between the first and the second image presented on each trial. The response for each trial was collected after both the images had been displayed and removed from the screen. The observer could take as long as they wished before entering the response. Feedback was provided via tones presented after the response to allow observers to maximize their performance. The next trial was presented 250ms (ITI) after the feedback. Thus, the actual inter-trial interval depended on the response time of the observer.*

**Psychometric Function**

The proportion comparison chosen data was used to obtain the psychometric function for each block. Each block consisted of 330 trials with 30 trials at each comparison lightness level. At each lightness level, we recorded the number of times the observers chose the comparison image to be lighter. The proportion comparison chosen data were fit with a cumulative normal using the Palamedes toolbox (Prins & Kingdom, 2018) to obtain four parameters of the psychometric function: threshold, slope, lapse rate and guess rate. The lapse rate was constrained to be equal to the guess rate and to be in the range [0, 0.05]. The psychometric function was fit using the maximum likelihood method. The threshold was obtained as the difference between the LRFs at proportion comparison chosen 0.76 and 0.50 as obtained from the cumulative normal fit.

**Ethics Statement**

All experimental procedures were approved by North Carolina Agricultural and Technical State University Institutional Review Board and were in accordance with the World Medical Association Declaration of Helsinki.

**Code and Data Availability**

*For each observer, the proportion comparison chosen data for the 18 experimental blocks as well as the thresholds are provided as supplementary information (SI). The SI also provides the MATLAB scripts to generate Figures 2, 4, 5, 6 and 7 and the scripts to obtain thresholds of the linear receptive field formulation of the model (model described below). The computed retinal images used as input to the model are provided as .mat files in a zip folder. The SI is available at: https://github.com/vijaysoophie/EquivalentNoisePaper.*

**3 MODEL**

*The data collected in the experiments characterize how lightness discrimination thresholds increase with the variance of a task-irrelevant stimulus variable. Interpreting the data is aided by a model that relates the changes in discrimination thresholds to the underlying precision of the perceptual representation. The model provides a way to connect the variance of a task-irrelevant property to the precision of the perceptual representation of the task-relevant stimulus variable (here lightness). The model we employ shares features of models that have been used to understand how contrast thresholds are elevated in the presence of contrast noise (see e.g., Legge, Kersten, & Burgess, 1987; Pelli, 1990). We provide a full development of the model here, however, as the current application of the underlying ideas differs substantially from previous applications.*

*We first introduce an analytic formulation, derived in the context of signal detection theory (SDT formulation). We then show how this can be instantiated as a linear receptive field model whose performance can be simulated (LINRF formulation). An important advantage of the LINRF formulation is that it can accommodate the physical-realizability constraint incorporated into our statistical model of naturally occurring reflectances.*

*The model allows us to express the variation of the task-irrelevant stimulus variable in units of equivalent noise standard deviation, where an equivalent noise standard deviation of 1.0 corresponds to the amount of external variation whose effect on the perceptual representation of the task-relevant stimulus variable is the same as that of the intrinsic internal noise that limits discrimination in the absence of task-irrelevant external variation. In this way, we can understand the effect of the task-irrelevant variability on thresholds in perceptually meaningful units of equivalent noise level. Task-irrelevant variability with an equivalent noise level less than one have little impact on the visual system, since its effects are dominated by intrinsic variability. Levels of task-irrelevant variability with an equivalent noise level greater than one do intrude on perception. The equivalent noise level indicates the magnitude of the intrusion in units that connect to intrinsic precision. Equivalent noise is similarly used in the literature on contrast noise masking (again see e.g., Legge, Kersten, & Burgess, 1987; Pelli, 1990).*

***SDT Model Formulation***

*We first formulate our model in the context of signal detection theory (Green, 1966). We model the visual response to the target object in each image by a univariate internal representation denoted by the variable . This variable depends on the image and is perturbed by noise. We assume that for any fixed image, is a normally-distributed random variable whose mean depends on the target object LRF. For each image, we assume that is perturbed on a trial-by-trial basis by independent zero-mean normally-distributed noise, and we assume that the variance of this noise is the same for the response to all images. We refer to the noise that perturbs for a fixed image as the internal noise and denote its variance as For each trial of the experiment, takes on two values, and , one for the interval containing the standard and the other for the interval containing the comparison.*

*If we consider performance for a particular pair of target standard and comparison LRFs, performance depends both on the difference between the expected values of for each pair of LRFs, and , and on the value of In our experimental design we have ensembles of images with different backgrounds, for each value of the target object LRF and background covariance scalar. The fact that we draw stochastically from these ensembles on each trial introduces additional variability into the value of the decision variable that corresponds to a fixed target LRF. We call this the external variability, and model it as a normal random variable with zero mean and variance We assume that depends on the experimentally chosen covariance scalar, but not on the target sphere LRF. Thus, the distributions of and , for a particular choice of target standard and comparison LRF and covariance scalar, are given by and . Here is the mean value of the internal representation to the standard image and is the mean value of the internal representation to the comparison image. The overall standard deviation is obtained via , where and are the variance of the internal and external noise.*

*For a 2AFC discrimination task in the context of signal detection theory, the observer makes their decision based on a comparison of and , choosing the interval with the higher value of as that with the higher stimulus value. The observer’s sensitivity depends on the mean values and the variance of , and is captured by the quantity d-prime: . D-prime measures the distance between the two distributions in standard deviation units. A value of corresponds to an inability to distinguish between the standard and the comparison image. Larger values of indicate increasing discriminability.*

*For a fixed value of , the difference in mean values is directly proportional to the standard deviation :*

*. (1)*

*We further assume that the difference in mean value of the internal variable is proportional to the difference in the LRFs of the target object in the standard and comparison images (). That is, , where is the proportionality constant. This yields*

*(2)*

*When we measure threshold in a 2AFC task, we choose a criterion proportional correct and find the that corresponds to that proportion correct. Our choice of 0.76 corresponds to In addition we can choose , in essence setting the units for to match those of the target LRF.*

*In our experiment, external variability was induced by changing the reflectance of the objects in the background. We used a multivariate normal distribution to generate the reflectance functions of the background objects.[[4]](#footnote-4) To change the amount of external noise, we scaled the covariance of the multivariate normal distribution by multiplying its covariance matrix with a scalar. Thus, for our experiments we have*

*(3)*

*where is the covariance scalar and is the external noise introduced when the ensemble of images for each value of target LRF has the reflectance of the background objects drawn from our model of natural reflectances.*

*Converting the equation above to the form we use to represent the data, we have*

*. (4)*

*The equation above predicts that the form of threshold* *as a function of covariance scalar should increase monotonically. For small values of (/), the threshold will approach a constant giving For large values of (/), the quantity will approach a straight line with slope 1 in the versus plot. Fitting the measurements with Equation 4 allows us to check whether the model describes the data, as well as to determine the two parameters and . In particular, we can establish the relative contribution of the internal representational variability and external stimulus variability in limiting lightness discrimination. The parameter quantifies how much the variation in background object reflectances intrudes on the internal representation that mediates the lightness discrimination task. The value of may be compared directly to the intrinsic precision of that representation characterized by*

***Equivalent Noise Level***

*The SDT formulation allows us to introduce the concepts of equivalent noise and equivalent noise level. The equivalent noise is the amount of external variation that has the same effect on the decision variable as the internal noise. The external variation is characterized experimentally by the covariance scalar (together with the underlying model of natural reflectances which is held fixed across the experiments). Once the model parameters and are determined from the data, we can find the covariance scalar that produces externally-generated equivalent noise*

*(5)*

*This in turn allows us to express the covariance scalars in terms of their equivalent noise level, which gives their effect on the perceptual representation relative to the effect of the internal noise. Thus*

*. (6)*

*For , the effect of the external noise is negligible and does not affect the perceptual representation and the internal noise dominates the precision of the representation. For , the effect of the external noise dominates the perceptual representation, and the visual system has not insulated the representation of the task-relevant stimulus variable from the variation in the task-irrelevant perceptual variable. When the equivalent noise level is ~1, the effect of the external variability is matched to that of the internal variability. At this operating point, further insulation of the task-relevant representation will not lead to significant further increases in the precision of this representation. We can thus use the equivalent noise level as a calibrated metric for assessing the magnitude of the perceptual effect of various levels of task-irrelevant stimulus variation.*

***Linear Receptive Field Formulation***

*When external noise added to the images is characterized by a multivariate normal and the decision noise is normal, a simple linear receptive field (LINRF) formulation is equivalent to the SDT formulation developed above. We develop this equivalence below. The advantage of the LINRF formulation is that it can easily be applied directly to images and to cases where the internal or external variability is non-normal. In our application, there are two non normalities. First, although the projection weights for linear model of naturally-occurring reflectance are drawn from a multivariate normal distribution, the constraint that the resulting reflectance functions lie within the range between 0 and 1, implemented to satisfy physical realizability, makes the overall distribution non-normal. Second, we incorporate into the model the Poisson variability of the cone excitations.*

*We begin with development that connects the LINRF formulation to the SDT formulation. In the LINRF formulation, the decision variable is computed from the displayed stimulus as the response of a single unit whose responses are a linear function of the stimulus image. Denote the stimulus image by the column vector , and the receptive field by the column vector . The entries of are the radiant power emitted by the monitor at each image location. The entries of are the corresponding sensitivities of the linear receptive field to each entry of . The response of the receptive field is given as , where is a random variable representing a draw of zero mean normally-distributed internal noise (variance ) in the receptive field response for a fixed image. We assume that is independent of .*

*Denote and as the standard and comparison images without external noise. External normally-distributed noise is added to both and , with covariance matrix . The external noise need not have zero mean. After incorporation of the external noise, the response of the receptive field to the comparison and standard images is given by*

*(7)*

*. (8)*

*Here is a random variable representing a draw of external noise, represents the internal noise, and is a random variable representing the overall effect of the external and internal noise. Since the receptive field and noise models are linear and normal, is normal with variance*

*(9)*

*The mean difference between the receptive field response to the comparison and the standard image is given by . Here and are the standard and comparison images without external noise added is a constant, and is as defined is the SDT section above. The second equality follows because 1) the difference between and is proportional to as only the target LRF changes between these two images and 2) even if the mean of the external noise is non-zero, its effect cancels when we obtain the mean difference in response.*

*We associate the linear receptive field response with the internal representation of the SDT formulation developed above. That is, we assume that on each trial, the observer chooses as lighter the interval for which the response of the receptive field is greater. Following the development of the SDT formulation, we have*

*(10)*

*where we have introduced the covariance scalar in the term corresponding to the variance of the external noise, and where denotes the covariance matrix of the external noise corresponding to the level of variation in natural images. Comparing to relation derived in the SDT model (Equation 3), we see that this is the same functional form for the relation between and as derived there, where we associate and .*

*To fit the LINRF formulation and relax its assumptions, we compute how images produce retinal cone excitations and employ a one-parameter description of a simple center-surround receptive field that draws upon the output of the cones. We use simulation to compute model responses for any choice of . This procedure is described in more detail below. Once the fitting procedure establishes and that best account for the data, we then find directly by passing the images corresponding to through the receptive field and finding the resulting variance. These parameters in turn allow us to compute and for the LINRF formulation.*

***Fitting the SDT Model Formulation***

*The model**was**fit to the threshold versus covariance scalar data to obtain the parameters and . The parameters were obtained by minimizing the mean squared error between the measured and predicted threshold using the MATLAB function fmincon. The best fitting parameters were estimated separately for the mean observer and the individual observers.*

***Fitting the Linear Receptive Field Model Formulation***

*We fit the linear receptive field (LINRF) model using a simulation approach. We used simulation for two reasons. First, it allows us to incorporate a model of the early visual system into the computations. Second, it provides a way to account for truncation in the normally-distributed model of natural reflectances.*

*The model of initial visual encoding was as described by Singh et al. (2018), and was implemented using the software infrastructure provided by ISETBio (ISETBio; isetbio.org; Cottaris, Jiang, Ding, Wandell, & Brainard, 2019). It incorporated typical optical blur (Thibos, Hong, Bradley, & Cheng, 2002) and the Poisson noise that perturbs cone photoreceptor isomerizations in the retina (Rodieck, 1998). In addition, it included axial chromatic aberration (Marimont & Wandell, 1994), and spatial sampling by the mosaic of long (L), middle (M) and short (S) wavelength-sensitive cones (Brainard, 2015). The L:M:S cone ratio in the cone mosaic was chosen to be 0.6:0.3:0.1 (1523 L-cones, 801 M-cones, 277 S-cones). The CIE physiological standard (CIE, 2007), as implemented in ISETBio, was used to obtain LMS cone fundamentals. Cone excitations were calculated as the number of photopigment isomerizations in a 100ms integration time, and included simulation of the Poisson variability of the isomerizations (Rodieck, 1998). The cone isomerizations were demosaiced using linear interpolation to estimate LMS isomerization images. Further, the isomerizations of each cone class was normalized by the summed (over wavelength) quantal efficiency of the corresponding cone class, to make the magnitude of the signals from the three cone classes similar to each other. This normalization occurred after incorporation of Poisson noise and did not affect the signal-to-noise ratio of the signals from the different cone classes.*

*The dot product of the LMS isomerization images was taken with a simple center-surround linear receptive field. The receptive field was square in shape to match the image size. Its center was a circle of radius equal to the size and at the location of the target object in the image. The central region was taken to have spatially-uniform positive sensitivity, while the surround was taken to have spatially-uniform negative sensitivity. Each point in the central region had sensitivity and each region of the surround had sensitivity denoted by . The RF was the same for each of the three cone classes. The RF response was taken as the sum of the L, M and S RF component responses. Normally-distributed internal noise with zero mean was added to the resulting dot product. The variance of the internal noise () and the value of the RF surround sensitivity () were the two parameters of the model.*

*The threshold predictions of the LINRF formulation for any choice of model parameters were obtained using simulation of a two-interval force choice paradigm similar to the experiment. For each trial, we randomly sampled a standard image and a comparison image from our dataset, following the procedure used in the experiment. We obtained the response of the receptive field (noise-added dot product) to the images and compared them to determine the simulated choice on that trial. This process was repeated 10,000 times for each of the 11 comparison LRF levels. The proportion comparison chosen data were used to fit the psychometric function and obtain the discrimination threshold, similar to the method used for the human psychophysical data. We estimated model threshold for the six values of covariance scalar at which we performed the human experiments.*

*We calculated the mean squared error (averaged over the six covariance scalar values) between the thresholds of the human data being fit and the computational model for a large set of values of the two model parameters: the variance of the decision noise () and the value of the RF surround (). The mean squared error values obtained as a function of these two parameters were fit with a degree two polynomial of two variables using the MATLAB fit function. The resulting polynomial was evaluated to estimate the parameters with lowest mean square error. These parameters were then used to estimate the internal and external noise standard deviation of the LINRF formulation using the relations: and as explained above, where the constant was obtained by solving .*

*The best fitting parameters were estimated separately for the mean observer and the individual observers.*

**4 RESULTS**

**Human Lightness Discrimination Thresholds Increase with Background Reflectance Variation**

We measured lightness discrimination thresholds of human observers for two types of variation 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 thus the background objects varied in their chromaticity and luminance. In achromatic variation, each spectrum had the same reflectance at all wavelengths, and thus the spectra varied only in their overall luminance and the objects were gray in color. The amount of variation depended on the covariance matrix of the multivariate normal distribution from which the spectra were sampled. The variance was controlled by multiplying the covariance matrix by a covariance scalar (). We measured discrimination thresholds of six human observers at six values of the covariance scalar for chromatic variation and three values of covariance scalar for achromatic variation. The threshold was measured three times (three separate blocks) for each observer and each of the nine conditions. The psychometric functions for these nine conditions are shown for one observer in Figure 4 and for all observers in Figure S3. Inspection of the psychometric functions show that their slopes steadily decrease with increasing covariance scalar, corresponding to an increase in thresholds. The thresholds for chromatic and achromatic variation are comparable.

Figures 5 shows explicitly how the discrimination thresholds change with the amount of variability in the reflectance of the background objects. Here, we plot the mean log threshold squared (averaged across observers, N = 6) against the log of the covariance scalar. Table S1 provides the thresholds and SEMs from Figure 5 in tabular form. For low values of the covariance scalar, the thresholds are nearly constant and are similar across observers. As the covariance scalar increases, log squared threshold increases. The thresholds are comparable for chromatic and achromatic variation. These features are seen in the mean data (Figure 5) and in the data for all observers (Figure 6).

[MAKE LOG THREHSOLD FIGURE FOR INDIVIDUAL OBSERVERS]

**Equivalent noise characterization of background variation**

**Human Lightness Discrimination Thresholds Increase with Light Source Intensity Variation**

We measured lightness discrimination thresholds of human observers as we varied the intensity of light sources in the scene. The spectrum of light sources was fixed to be standard daylight spectrum D65. We normalized the spectrum by its mean over wavelengths. The intensity was varied multiplying the normalized spectrum by a scalar sampled from a log-uniform distribution in the range [1- \zeta, 1+ \ zeta]. The reflectance spectra of the background objects were fixed. We measured lightness discrimination thresholds for seven values of the range parameter \zeta for five human observers. The psychometric function of one of the observers for these seven conditions are shown in Figure XX. Figure XX shows the mean threshold of the five observers. Similar to the trend for reflectance spectra variation, lightness discrimination thresholds remain constant for small values of the range parameter and then log threshold squared increases with increase in range parameter. A fit of the mean threshold with the linear receptive field model gives the value of internal noise as XXX. This compares well with the internal noise obtained using the thresholds obtained from background reflectance spectra variation.

**Thresholds for Simultaneous Variation are Higher Than Individual Variations**

Finally, we measured lightness discrimination thresholds for simultaneous variation in the reflectance spectra of background objects and the intensity of the light sources in the scene. In this experiment, we studied six conditions: no variation, achromatic and chromatic variation in the background objects with fixed light source spectrum, variation in intensity of light source with fixed background, and simultaneous variation in the intensity of light source and background object reflectance spectra for chromatic and achromatic. We measured lightness discrimination thresholds of six human observers for these six conditions. The psychometric function of one of the observers is shown in Figure XX. Figure XX shows the psychometric functions of all observers. Figure XX shows the mean threshold of all six observers for these six conditions. We see that the threshold for simultaneous variation of light intensity and reflectance spectra of background objects is higher than the condition with individual variations. As observed earlier, the threshold for achromatic and chromatic conditions are comparable.

Figure XX shows the increase in mean squared threshold above the no variation condition. We compare the mean square thresholds of the simultaneous variation condition with the sum of the mean square thresholds of the individual conditions for chromatic and achromatic conditions. The increase in threshold of the simultaneous variation condition is comparable to the sum of the increase in threshold for the individual variations.

We used the linear receptive field parameters obtained from the background reflectance variation condition on the images of this experiment. Figure XXX shows the thresholds of the linear receptive model for the six conditions. As expected, the threshold of the linear receptive model are within comparable to the measured threshold of the no-variation condition and background spectra variation conditions. Also, the threshold of the linear receptive model is significantly higher than the measured threshold of the light source intensity variation condition. Surprisingly, the threshold of the linear receptive field model for the simultaneous variation condition are comparable to the measured threshold for this condition.

**5 DISCUSSION**

*The perceived**lightness**of an object can depend on the scene in which it lies. Stabilization of the lightness representation against variation in scene properties extrinsic to the object’s surface reflectance is referred to as lightness constancy. In this paper, we introduced a new psychophysical approach for characterizing lightness constancy. The approach is based on measuring how lightness discrimination thresholds vary with experimentally introduced variation in scene properties extrinsic to the object’s reflectance. Specifically, we studied how lightness discrimination thresholds are impacted by variation in the reflectance of the background objects in naturalistic scenes rendered using computer graphics. Our results (Figures 5 and 6) show that when the variation in the reflectance of background objects is small, discrimination thresholds are nearly constant. In this regime, performance is limited primarily by internal noise. As the amount of background object reflectance variation increases, the effect of external variation starts dominating that of the internal noise, and discrimination thresholds increase. We analyzed the data using a modeling approach used previously to study effect of external noise on contrast detection (Legge, Kersten, & Burgess, 1987; Pelli, 1990; Pelli & Farell, 1999). This approach allows us to relate the effect of background object reflectance variation to the intrinsic precision of the lightness representation. The intrinsic precision depends on the observer’s internal noise, which limits performance in the absence of external variation. The model compares discrimination thresholds with and without extrinsic variations to quantify variance in the perceptual representation of lightness induced by extrinsic variation. It allows us to express the effect of extrinsic variation as an equivalent noise level (, that is relative to the standard deviation of the intrinsic noise. In this way, we use the intrinsic noise as a benchmark to interpret the magnitude of the equivalent noise from the external variation. We find that the effect of the external variability introduced by variation of background object reflectances in naturalistic scenes is within a factor of two of the intrinsic precision of the lightness representation. More generally, our work provides a method to quantify the effect of variation in a task-irrelevant properties on the perception of task-relevant property, and is thus applicable to understanding other perceptual constancies beyond the lightness constancy we focused on here.*

***Relation to Contrast Detection in Contrast Noise***

*As noted, our paradigm and model have conceptual roots in the literature on contrast detection in contrast noise. The concept of equivalent noise plays an important role in this literature (Legge, Kersten, & Burgess, 1987; Pelli, 1990 ; Pelli & Farell, 1999). However, there is an important difference between the way the ideas are applied to understand contrast detection and the way we have leveraged them here. In the contrast detection literature, detection in the absence of external noise is conceptualized as limited by two distinct factors. One factor is the internal variability in the observer’s representation of contrast. The other factor is the efficiency with which the observer’s decision processes makes use of the information provided by this representation, which is inferred through an ideal observer analysis applied to high external noise conditions, where effects of internal noise are swamped by those of the external noise (Pelli, 1990 ; Pelli & Farell, 1999). This separation is enabled when such an ideal observer calculation is available, and in practice is more straightforward when the stimulus being detected/discriminated and the external noise being added have commensurate units (e.g. contrast energy). In our work, the task-relevant and task-irrelevant stimulus variables vary along distinct dimensions of the stimulus space (e.g., affect distinct image locations). Currently we do not have in hand an ideal observer calculation that would allow us to compute the visual system’s efficiency in using the available information. Obtaining and integrating such a calculation would be of interest. Singh, Cottaris, Heasly, Brainard, & Burge (2018) provide a possible approach, but employing that approach would require measurements with a larger set of task-irrelevant variation (e.g., illumination as well as background) than available from the current data.*

***Spatial and Chromatic Properties of the Stimuli***

*We used small image patches in our study. The small size of the image patches is a notable difference between our stimuli and natural viewing. In this initial deployment of our paradigm, we thus focused on effects of background object reflectance variation that are nearby the test object. The observed effects may be mediated by relatively small populations of neurons. The use of small image patches is not a necessary requirement of our paradigm, which could be extended to larger images. Such extension could reveal additional effects not captured by the current experiments.*

*In addition to using small patches, we did not vary the spatial structure of the array of objects in the rendered scenes. Manipulating spatial structure, in addition to increasing image size, may provide a way to use our paradigm to measure the spatial tuning of the mechanism(s) mediating the background effect. This approach is loosely analogous to how manipulating the structure of contrast noise may be used to examine the tuning of mechanisms supporting the detection of contrast-defined targets (Henning, Hertz, & Hinton, 1981; Rovamo, Franssila, & Nasanen, 1992; Losada & Mullen, 1995; Nachmias, 1999; Rovamo, Raninen, & Donner, 1999).*

*Although we restricted our measurements to lightness discrimination thresholds, our variation of the reflectance properties of the background objects was not limited to variation in overall reflectance. The choice to introduce background object reflectance variation along more spectral dimensions (affecting e.g. background object hue and saturation) than used for target object variation was somewhat arbitrary – we could have restricted the background object reflectance variation to one dimension (e.g. overall scale of reflectance spectra) or studied discrimination of additional (e.g. chromatic) dimensions of target object variation. As with the case of spatial structure above, extending the measurements to a wider range of stimuli is of interest. Indeed, it may be possible to manipulate the chromatic structure of the variation in background object reflectances with the goal of understanding the chromatic tuning of the background object reflectance variation’s effect on the lightness discrimination thresholds, as well as on other target object discriminations. This would again be analogous to how noise-based approaches have been used to characterize chromatic tuning of mechanisms that support the detection of chromatically-defined contrast targets (Gegenfurtner & Kiper, 1992; Sankeralli & Mullen, 1997; Giulianini & Eskew, 1998; Monaci, Menegaz, Süsstrunk, & Knoblauch, 2004).*

***Link Between Thresholds and Suprathreshold Perceptual Judgments***

*The technique developed here probes the constancy of a perceptual representation of a task-relevant variable (e.g., perceived object lightness) by measuring how variation in a task-irrelevant scene variable (e.g., background object reflectances) elevates thresholds for detecting changes in the task-relevant variable. As with other threshold-based methods for approaching the stability of suprathreshold perceptual judgments (see Introduction), the extent to which the results may be used to predict the stability such judgments across changes in other scene variables is not known. Experiments that explore this link, perhaps by directly comparing results from the two paradigms with similar stimuli and the same set of observers, are of considerable interest. The results of such experiments might also be helpful in pointing the way to theory that would link results across the two paradigms; at present we do not have such theory in hand (but see Abrams, Hillis, & Brainard, 2007).*

*Previous authors have suggested that lightness constancy improves with increasing background “articulation”. That is, increasing the number of objects in the background and/or the degree to which their reflectance varies tends to improve constancy (Gilchrist, 2006; Radonjić & Gilchrist, 2013; see also Radonjić, Cottaris, & Brainard, 2015; Kraft, Maloney, & Brainard, 2002). This may on the surface seem in contradiction to our results; we find increasing the variance of the background reflectances has a deleterious effect on lightness discrimination performance. Note, however, that articulation is thought to improve constancy when the task-irrelevant variation is a change in illumination, and where the background itself is held fixed across this change. In our experiments, the illumination is held fixed and we consider the effect of the background per se, with the background change occurring across the two intervals of each forced-choice trial. Thus, we are studying a different aspect of lightness constancy than where increased articulation is thought to lead to improvements, and our results are not in conflict with previous findings.*

*Our paradigm could be used to study constancy across changes in illumination, if the task-irrelevant variation used in the experiment were in the illumination rather than the background object reflectances. In that case, the articulation idea would predict a smaller elevation of lightness discrimination thresholds when the effect of illumination variation was studied for scenes with higher variance in the background reflectance, as long as the background was held fixed across the two intervals of each trial.*

***Applications to Understanding Neural Mechanisms***

*A longstanding goal of vision science is to connect psychophysical performance to its underlying neural mechanisms. For probing mechanisms that mediate perceptual constancies, our paradigm has the attractive feature that there is a well-defined correct answer on each trial, so that for studies with animal subjects it is possible to provide performance-contingent reward. In addition, there are well-worked out methods for predicting psychophysical discrimination performance from recordings of the responses of neural populations (Shadlen, Britten, Newsome, & Movshon, 1996; Parker & Newsome, 1998; Cohen & Newsome, 2009; Nienborg, Cohen, & Cumming, 2012; Ruff, Ni, & Cohen, 2018), and the theoretical links between such analysis and performance should continue to hold when task-irrelevant stimulus variation is added to the paradigm. Complementing neural measurements that include random, unpredictable task-irrelevant stimulus variation with such analyses may provide rigorous quantitative insights about the sensory-perceptual processing and the neural computations underlying color and lightness constancy specifically, and perceptual constancy more generally.*

***Model of Natural Surface Reflectances***

*We used a truncated multivariate normal distribution as the statistical model for the projection weights of a linear model of naturally occurring reflectances, to sample the background object reflectance functions. This model was developed in our earlier work and is evaluated more fully there (Singh, Cottaris, Heasly, Brainard, & Burge, 2018; see also Brainard & Freeman, 1997; Zhang & Brainard, 2004). The model is based on measurements of the surface reflectance functions of the Munsell papers (Kelly, Gibson, & Nickerson, 1943) as well as natural surfaces characterized by Vrhel (1994). The underlying multivariate normal provides a convenient way to capture two basic aspects of natural variation in reflectance. First, these reflectances are well-described by low-dimensional linear models (Cohen, 1964; Maloney, 1986; Parkkinen, Hallikainen, & Jaaskelainen, 1989). Second, within the reflectance subspace defined by the linear models, not all reflectances are equally likely to occur. Still, we think it likely that future work will lead to more accurate statistical models of naturally occurring reflectance. For example, it is possible that replacing the linear model approach with a prior that favors spectrally-smooth reflectance functions (Jiang, Farrell, & Wandell, 2016) would lead to a more accurate characterization. In addition, we have assumed that the distribution of reflectance functions over objects is independent, but this assumption may not be accurate. Approaches to modeling a dependency have been suggested (Shen & Yeo, 2011; Gehler, Rother, Kiefel, Zhang, & Schölkopf, 2011; Barron & Malik, 2012b; Barron & Malik, 2012a).*

*It is important to note that the quantitative relation we measured between the magnitude of internal noise and the effect of external noise introduced as variation in background object reflectances depends on how the distribution of naturally-occurring reflectances is modeled. If the model of reflectances overestimates the natural variation, the effect of external noise in natural scenes will be less than we estimated. Conversely, if the model of reflectances underestimates the natural variation, the effect of external noise in natural scenes will be greater than we estimated. Importantly, improved future characterization of naturally occurring reflectances, obtained through the acquisition of additional reflectance measurements and advances in their statistical description, could be used in conjunction with the parameters of the LINRF model formulation, without need for new data collection, to update the estimate of the effect of naturally occurring background object reflectance variation on object lightness perception.*

***Rule of Combination***

*In the present work, we considered variation in only a single task-irrelevant variable. In natural scenes, there are many task-irrelevant variables. In the case of judging object lightness, these include object-extrinsic factors such as the scene illumination, the position and 3D orientation of the target object in the scene, the viewpoint from which the object is viewed, and various object-intrinsic factors like its shape and size. Variation in each of the factors could in principle elevate thresholds for discriminating object lightness. Our paradigm allows characterization of the effect of these task-irrelevant variables and quantifies that effect for each such variable in the same internal-noise referred units. One potentially important future direction is to measure the combined effect of simultaneous variation of multiple task-irrelevant variables, and to test hypotheses about rules of combination that predict the joint effects of such simultaneous variation.*

**6 ACKNOWLEDGEMENTS**: NSF BCS 2054900 (VS).

**APPENDIX**

**Measurement of object lightness discrimination thresholds under variation in background object reflectances**

The control experiment, preregistered as Experiment 2, provided preliminary data that helped shape the design of the main experiment presented in the paper (which was Experiment 3 of the preregistration documents). It aimed to determine whether variation in the reflectance of background objects had an effect on human lightness discrimination thresholds. It established that human object lightness discrimination thresholds increase if the reflectances of background objects vary, as compared to the case when the discrimination is made against a constant background. It also studied the effect of inclusion or not of secondary reflections in the rendering process and assessed the effect of implementing background object reflectance variation across trials rather than across intervals.

The basic methods were the same as for preregistered Experiment 3. The practice session was conducted with the images in Condition 1 described below. The observers were retained for the experiment if their average threshold of the last two blocks during the practice session was lower than 0.030. This was a deviation from the preregistered plan where we set the threshold criterion as 0.025. After collecting data from 8 observers, we realized that the criterion was too strict. Only one observer had met the criterion. After modifying the threshold criterion, we included two of the initially discontinued observers in our experiment (Observer 5 and Observer 8). A total of 11 naïve observers participated in the practice sessions. Four of these observers met the criteria for continuing the experiment. Two of these observers also participated in the main experiment (Observer 4 and Observer 8). The visual acuities of these 4 observers were: Observer 4, L = 20/15, R = 20/20; Observer 5, L = 20/20, R = 20/40; Observer 8, L = 20/30, R = 20/25; Observer 11, L = 20/25, R = 20/30. Observers 5, 8, and 11 wore personal corrective eyewear both during vision testing and during the experiments. Observer 4 did not require or use corrective eyewear.

We measured lightness discrimination threshold of four naïve human observers using a two-interval forced choice paradigm. The thresholds were measured for three specific types of background variation (Figure S1). The reflectance spectra of the background objects were generated with the covariance scalar set to 1. These three conditions were:

*Condition 1.* *Fixed background:* In this condition, the spectra of objects in the background were kept fixed for all trials and for all intervals. We generated 11 images, one at each comparison LRF level.

*Condition 2. Between-trial background variation*: In this condition, the spectra of the objects in the background were the same for the two intervals within a trial but varied from trial-to-trial.

*Condition 3. Within-trial background variation*: In this condition, the spectra of the objects in the background varied between trials as well as between the two intervals of a trial. The background variation corresponded to covariance scalar equal to 1.

In Conditions 2 and 3, the light reflected from the target object varied from image to image (even at the same LRF level of the target object) because of secondary reflection of light coming from the background objects was included in the rendering. We also measured the thresholds without secondary reflections for these two conditions. We call these conditions Condition 2a and 3a.

*Condition 2a. Between-trial background variation without secondary reflection*: Same as Condition 2, but without multiple reflections of light from object surfaces. The light rays only bounce off once from the surfaces before coming to the camera.

*Condition 3a. Within-trial background variation without secondary reflections*: Same as Condition 3, but without multiple reflections of light from object surfaces. Condition 3a was the same as the experiment reported in the main paper for covariance scalar equal to 1.

Figure S2 shows the discrimination thresholds of the four human observers for the five conditions studied in this experiment. We plot the mean threshold and the standard error of the mean (SEM) taken over the three separate threshold measurements. For each observer, the thresholds for Conditions 3 and 3a were higher compared to Conditions 1, 2 and 2a. The average increases in threshold of the observers for Conditions 3 and 3a as compared to Condition 1 (baseline) were 79% and 60% respectively. The average increases in threshold for Conditions 2 and 2a were much smaller, 13% and 17% respectively. The thresholds for Conditions 1, 2 and 2a were nearly within one SEM of each other (averaged over the observers and three conditions). On the other hand, the thresholds for Conditions 3 and 3a were respectively (on average) 7.2 and 5.4 SEM larger than the threshold of Condition 1. The thresholds without secondary reflections (Conditions 2a and 3a) were within one SEM from the conditions with secondary reflections (Conditions 2 and 3).

The control experiment established that lightness discrimination thresholds are higher for the case when the two objects are being discriminated against different backgrounds on the same trial, as compared to when the backgrounds are the same within trial. Trial-to-trial variability in background object reflectances across trials has little, if any, effect. The effect is similar when the rendering is performed with and without secondary reflections, indicating the effect is due to the spectral change in the background and not due to the variation in the amount of light being reflected from the target object. In the main experiment, we rendered without secondary reflections to avoid introducing such variability. Figure S2 also shows the threshold of the observers in the main experiment (preregistered Experiment 3) for the condition with covariance scalar equal to 1. This condition is equivalent to Condition 3a of the control experiment (preregistered Experiment 2). Thresholds were consistent across the two measurements.

**Table S1: Thresholds for Control Experiment (Preregistered Experiment 2):**Mean threshold (averaged over blocks) SEM of four human observers for five background variation conditions studied in experiment 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mean Threshold +- SEM (averaged over sessions) | | | | |
| Observer | Condition 1 | Condition 2 | Condition 2a | Condition 3 | Condition 3a |
| 4 | 0.0269+-0.0013 | 0.0254+-0.0013 | 0.0235+-0.0011 | 0.0366+-0.0030 | 0.0330+-0.0018 |
| 5 | 0.0217+-0.0005 | 0.0305+-0.0039 | 0.0300+-0.0017 | 0.0382+-0.0031 | 0.0389+-0.0022 |
| 8 | 0.0167+-0.0011 | 0.0169+-0.0020 | 0.0175+-0.0017 | 0.0325+-0.0016 | 0.0273+-0.0016 |
| 11 | 0.0252+-0.0013 | 0.0268+-0.0018 | 0.0285+-0.0002 | 0.0525+-0.0038 | 0.0439+-0.0068 |

**Table S2. Thresholds for Main Experiment (Preregistered Experiment 3)**:  
Mean threshold (averaged over blocks) SEM of four human observers measured at six logarithmically spaced values of the covariance scalar.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Observer | Covariance Scalar | | | | | |
| 0 | 0.01 | 0.03 | 0.1 | 0.3 | 1 |
| 2 | 0.0217+-0.0009 | 0.0238+-0.0006 | 0.0307+-0.0036 | 0.0294+-0.0008 | 0.0392+-0.0005 | 0.0429+-0.0049 |
| 4 | 0.0241+-0.0035 | 0.0215+-0.0015 | 0.0271+-0.0019 | 0.0246+-0.0018 | 0.0299+-0.0020 | 0.0295+-0.0014 |
| 8 | 0.0266+-0.0019 | 0.0214+-0.0005 | 0.0221+-0.0008 | 0.0273+-0.0024 | 0.0269+-0.0020 | 0.0318+-0.0041 |
| 17 | 0.0224+-0.0020 | 0.0236+-0.0030 | 0.0315+-0.0024 | 0.0347+-0.0027 | 0.0390+-0.0046 | 0.0454+-0.0032 |

**Figure Captions**

**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 spectra 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 reflectances. **(b)** Trial sequence. RN-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 RN 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]. In each block, 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 constrained 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 0.76 and 0.5, as predicted by the cumulative normal fit. This figure shows the data for Observer 2 for scale factor 0.00, for the block run 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 object reflectances:** The reflectance spectra of background objects 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 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 reflectances. 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.

**Figure 4: Psychometric functions for Observer 2.** We measured the proportion comparison chosen data at six values of the covariance scalar (), separately in three blocks for each observer. 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 for each of the three blocks, for Observer 2. The values in the legend provide the estimate of lightness discrimination threshold for each block 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 fit of the STD formulation of the model (Equation 4) is shown as the red curve. The parameters corresponding to this fit are provided in the legend. The threshold of the fit linear receptive field (LINRF) formulation was estimated by simulation 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 where and , , and are parameters adjusted in the fit. This functional form was chosen simply to provide a smooth curve through the simulated thresholds and has no theoretical significance. The parameters of the LINRF fit are also provided in the legend.

**Figure 6: Threshold of individual human observers.** Mean (across sessions) squared threshold vs log covariance scalar for individual human observers. Same format as Figure 5; here the error bars represent +/- 1 SEM taken across the three blocks for each observer. The parameters of the SDT and LINRF formulations were obtained separately for each observer and are provided in the legend, in order

**Figure 7. Equivalent noise analysis. (a)** The left panel shows the parameter estimates for the two model formulations for the mean data and each individual observer. From these, we can estimate the equivalent noise level (for background object reflectance variation corresponding to the full model of natural reflectance variation (covariance scalar **(b)** The equivalent noise level is provided for the mean data and each individual observer in the right panel.

**Figure S1: Control experiment stimuli.** Example stimuli for Conditions 1, 2 and 3 in the control experiment (preregistered Experiment 2) to study the effect of variation in background object reflectances on lightness discrimination threshold. In condition 1, the background was fixed in every trial and every interval. In Condition 2, the background object reflectances varied from trial to trial, but remained fixed in the two intervals of a trial. In Condition 3, the background object reflectances 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 Conditions 2 and 3 respectively, but without secondary reflections.

**Figure S2:** **Control experiment.** Lightness discrimination threshold of four human observers in the five conditions in the control experiment (preregistered Experiment 2). The plotted points have been jittered horizontally to avoid marker overlap. The thresholds are higher for the condition where the target objects are compared against a change in background object reflectances (Conditions 3 and 3a) than when the background is held fixed within each trial (Conditions 1, 2, 2a). Secondary reflections do not have any significant effect on thresholds (Conditions 2a and 3a). Condition 3a of the control experiment is equivalent to the condition of the main experiment (preregistered Experiment 3) with covariance scalar equal to 1. The thresholds for this condition of the main experiment are plotted here for comparison (). Two observers from the control experiment also participated in the main experiment.

**Figure S3: Psychometric functions for all observers.** Same asFigure 4 for all observers retained in the main experiment.

**REFERENCES**

Abrams, A. B., Hillis, J. M., & Brainard, D. H. (2007). The relation between color discrimination and color constancy: when is optimal adaptation task dependent? *Neural Computation, 19*, 2610-2637.

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science, 262(December 24)*, 2042-2044.

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences, 2nd edition* (pp. 339-351). Cambridge, MA: MIT Press.

Afifi, M., Barron, J. T., LeGendre, C., Tsai, Y.-T., & Bleibel, F. (2021). *Cross-camera convolutional color constancy*. Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision,

Allred, S. R., & Brainard, D. H. (2013). A Bayesian model of lightness perception that incorporates spatial variation in the illumination. *Journal of Vision, 13(7)*, 1–18.

Alvaro, L., Linhares, J. M. M., Moreira, H., Lillo, J., & Nascimento, S. M. C. (2017). Robust colour constancy in red-green dichromats. *PLoS ONE, 12(6)*, e0180310.

American Society for Testing and Materials. (2017). Standard test method for luminous reflectance factor of acoustical materials by use of integrating-sphere reflectometers. *Renovations of Center for Historic Preservation, 98(A)*, E1477.

Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal Of The Optical Society Of America A, 3(10)*, 1743-1751.

Aston, S., Radonjić, A., Brainard, D. H., & Hurlbert, A. C. (2019). Illumination discrimination for chromatically biased illuminations: implications for colour constancy. *Journal of Vision, 19(30:15)*

Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research, 27(11)*, 1915-1924.

Barron, J. T., & Malik, J. (2012a). *Color constancy, intrinsic images, and shape estimation.* Paper presented at ECCV.

Barron, J. T., & Malik, J. (2012b). *Shape, albedo, and illumination from a single image of an unknown object.* Paper presented at IEEE Conference on Computer Vision and Pattern Recognition, 334-341.

Bloj, M., Ripamonti, C., Mitha, K., Greenwald, S., Hauck, R., & Brainard, D. H. (2004). An equivalent illuminant model for the effect of surface slant on perceived lightness. *Journal of Vision, 4(9)*, 735-746.

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision, 3(8)*, 541-553.

Brainard, D. H. (1989). Calibration of a computer controlled color monitor. *Color Research & Application, 14(1)*, 23-34.

Brainard, D. H. (1998). Color constancy in the nearly natural image. 2. achromatic loci. *Journal of the Optical Society of America A, 15(2)*, 307-325.

Brainard, D. H. (2015). Color and the cone mosaic. *Annual Review of Vision Science, 1*, 519-546.

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image. 1. asymmetric matches. *Journal of the Optical Society of America A, 14(9)*, 2091-2110.

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical Society of America A, 14(7)*, 1393-1411.

Brainard, D. H., Longere, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M., & Xiao, B. (2006). Bayesian model of human color constancy. *Journal of Vision, 6(11)*, 1267-1281.

Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination models. *Journal of Vision, 11(5)*

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. In J. P. Hornak (Ed.), *Encylopedia of Imaging Science and Technology* (pp. 172-188). New York: Wiley.

Brainard, D. H., & Radonjić, A. (2014). Color constancy. *The New Visual Neurosciences, 1*, 545–556.

Brascamp, J. W., & Shevell, S. K. (2021). The certainty of ambiguity in visual neural representations. *Annual Review of Vision Science, in press*

Brindley, G. S. (1960). *Physiology of the Retina and the Visual Pathway*. London: Arnold.

Brown, R. O., & MacLeod, D. I. A. (1997). Color appearance depends on the variance of surround colors. *Current Biology, 7*, 844-849.

Burge, J. (2020). Image-computable ideal observers for tasks with natural stimuli. *Annual Review of Neuroscience, 6*, 491-517.

Burge, J., & Geisler, W. S. (2011). Optimal defocus estimation in individual natural images. *Proceedings of the National Academy of Sciences, 108(40)*, 16849-16854.

Burge, J., & Geisler, W. S. (2014). Optimal disparity estimation in natural stereo images. *Journal of Vision, 14(2)*

Burge, J., & Geisler, W. S. (2015). Optimal speed estimation in natural image movies predicts human performance. *Nature Communications, 6*, 7900.

Burge, J., & Jaini, P. (2017). Accuracy maximization analysis for sensory-perceptual tasks: computational improvements, filter robustness, and coding advantages for scaled additive noise. *PLoS Computational Biology, 13(2)*, e1005281.

Burnham, R. W., Evans, R. M., & Newhall, S. M. (1952). Influence on color perception of adaptation to illumination. *Journal of the Optical Society of America, 42(9)*, 597-605.

Chichilnisky, E. J., & Wandell, B. A. (1997). Increment-decrement asymmetry in adaptation. *Vision Research, 37*, 616.

Chin, B. M., & Burge, J. (2020). Predicting the partition of behavioral variability in speed perception with naturalistic stimuli. *Journal of Neuroscience, 40(4)*, 864-879.

CIE. (2007). *Fundamental chromaticity diagram with physiological axes – Parts 1 and 2. Technical Report 170-1*. Vienna: Central Bureau of the Commission Internationale de l' Éclairage.

Cohen, J. (1964). Dependency of the spectral reflectance curves of the Munsell color chips. *Psychon. Sci, 1*, 369-370.

Cohen, M. R., & Newsome, W. T. (2009). Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *J Neurosci, 29(20)*, 6635-6648.

Cottaris, N. P., Jiang, H., Ding, X., Wandell, B. A., & Brainard, D. H. (2019). A computational-observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-mosaic structure, and inference engine. *Journal of Vision, 19(4)*, 8.

Fechner, G. T. (1860). *Elements of Psychophysics* (H. E. Adler, 1966, Trans.). New York: Holt, Rinehart and Winston.

Flachot, A., & Gegenfurtner, K. R. (2018). Processing of chromatic information in a deep convolutional neural network. *JOSA A, 35(4)*, B334-B346.

Flachot, A., & Gegenfurtner, K. R. (2021). Color for object recognition: Hue and chroma sensitivity in the deep features of convolutional neural networks. *Vision Research, 182*, 89-100.

Foster, D. H. (2011). Color constancy. *Vision Research, 51(7)*, 674-700.

Gegenfurtner, K., & Kiper, D. C. (1992). Contrast detection in luminance and chromatic noise. *Journal of the Optical Society of America A, 9(11)*, 1880-1888.

Gehler, P., Rother, C., Kiefel, M., Zhang, L., & Schölkopf, B. (2011). *Recovering intrinsic images with a global sparsity prior on reflectance.* Paper presented at Advances in Neural Information Processing Systems, 765-773.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology, 59*, 167-192.

Geisler, W. S., Najemnik, J., & Ing, A. D. (2009). Optimal stimulus encoders for natural tasks. *Journal of Vision, 9(13)*, 17 11-16.

Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science, 195*, 185.

Gilchrist, A. L. (2006). *Seeing Black and White*. Oxford: Oxford University Press.

Giulianini, F., & Eskew, R. T., Jr. (1998). Chromatic masking in the (DL/L, DM/M) plane of cone-contrast space reveals only two detection mechanisms. *Vision Research, 38*, 3913-3926.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Vol. 1). New York: Wiley.

Heasly, B. S., Cottaris, N. P., Lichtman, D. P., Xiao, B., & Brainard, D. H. (2014). RenderToolbox3: MATLAB tools that facilitate physically based stimulus rendering for vision research. *Journal of Vision, 14(2)*

Helmholtz, H. (1896). *Physiological Optics*. New York: Dover Publications, Inc.

Helson, H., & Jeffers, V. B. (1940). Fundamental problems in color vision. II. Hue, lightness, and saturation of selective samples in chromatic illumination. *Journal of Experimental Psychology, 26(1)*, 1-27.

Helson, H., & Michels, W. C. (1948). The effect of chromatic adaptation on achromaticity. *Journal of the Optical Society of America, 38*, 1025-1032.

Henning, G. B., Hertz, B. G., & Hinton, J. L. (1981). Effects of different hypothetical detection mechanisms on the shape of spatial-frequency filters inferred from masking experiments: I. Noise masks. *Journal of the Optical Society of America, 71(5)*, 574-581.

Hillis, J. M., & Brainard, D. H. (2005). Do common mechanisms of adaptation mediate color discrimination and appearance? Uniform backgrounds. *Journal of the Optical Society of America A, 22(10)*, 2090-2106.

Hillis, J. M., & Brainard, D. H. (2007a). Distinct mechanisms mediate visual detection and identification. *Current Biology, 17(19)*, 1714-1719.

Hillis, J. M., & Brainard, D. H. (2007b). Do common mechanisms of adaptation mediate color discrimination and appearance? Contrast adaptation. *Journal of the Optical Society of America A, 24(8)*, 2122-2133.

Horn, B. K. P. (1974). Determining lightness from an image. *Computer Vision, Graphics, and Image Processing, 3*, 277-299.

Hurlbert, A. (2019). Challenges to color constancy in a contemporary light. *Current Opinion in Behavioral Sciences, 30*:186, 186-193.

Ishihara, S. (1977). Tests for colour-blindness. *Tokyo: Kanehara Shuppen Company, Ltd.*

Jaini, P., & Burge, J. (2017). Linking normative models of natural tasks to descriptive models of neural response. *Journal of Vision, 17(12)*, 16.

Jakob, W. (2010). Mitsuba Renderer.

Jameson, D., & Hurvich, L. M. (1955). Some quantitative aspects of an opponent-colors theory. I. Chromatic responses and spectral saturation. *Journal of the Optical Society of America, 45*, 546-552.

Jiang, H., Farrell, J., & Wandell, B. (2016). *A spectral estimation theory for color appearance matching*. Presented at the IS&T International Symposium on Electronic Imaging,

Kelly, K. L., Gibson, K. S., & Nickerson, D. (1943). Tristimulus specification of the Munsell book of color from spectrophoto-metric measurements. *Journal of the Optical Society of America, 33(7)*, 355-376.

Kingdom, F. A. (2011). Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research, 51(7)*, 652-673.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.

Kraft, J. M., Maloney, S. I., & Brainard, D. H. (2002). Surface-illuminant ambiguity and color constancy: effects of scene complexity and depth cues. *Perception, 31*, 247-263.

Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America, 61(1)*, 1-11.

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A, 4(2)*, 391-404.

Losada, M. A., & Mullen, K. T. (1995). Color and luminance spatial tuning estimated by noise masking in the absence of off-frequency looking. *Journal of the Optical Society of America A, 12(2)*, 250-260.

Lotto, R. B., & Purves, D. (1999). The effects of color on brightness. *Nature Neuroscience, 2(11)*, 1010-1014.

Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal Of The Optical Society Of America A, 3(10)*, 1673-1683.

Maloney, L. T., & Yang, J. N. (2001). The illuminant estimation hypothesis and surface color perception. In R. Mausfeld & D. Heyer (Eds.), *Colour Perception: From Light to Object* (pp. 335–358). Oxford: Oxford University Press.

Marimont, D. H., & Wandell, B. A. (1994). Matching color images: the effects of axial chromatic aberration. *Journal of the Optical Society of America A, 11(12)*, 3113-3122.

Monaci, G., Menegaz, G., Süsstrunk, S., & Knoblauch, K. (2004). Chromatic contrast detection in spatial chromatic noise. *Visual Neuroscience, 21*, 291-294.

Morimoto, T., & Smithson, H. E. (2018). Discrimination of spectral reflectance under environmental illumination. *J Opt Soc Am A Opt Image Sci Vis, 35(4)*, B244-B255.

Murray, R. F. (2020). A model of lightness perception guided by probabilistic assumptions about lighting and reflectance. *J Vis, 20(7)*, 28.

Murray, R. F. (2021). Lightness perception in complex scenes. *Annual Review of Vision Science, 7*, 417-436.

Nachmias, J. (1999). How is a grating detected on a narrowband noise masker? *Vision Research, 39(6)*, 1133-1142.

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: discrimination may be better than detection. *Vision Research, 14(10)*, 1039–1042.

Nienborg, H., Cohen, M. R., & Cumming, B. G. (2012). Decision-related activity in sensory neurons: correlations among neurons and with behavior. *Annu Rev Neurosci, 35*, 463-483.

Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. T. (2010). Categorical color constancy for real surfaces

. *Journal of Vision, 10(9)*

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annual Review of Neuroscience, 21(1)*, 227-277.

Parkkinen, J. P. S., Hallikainen, J., & Jaaskelainen, T. (1989). Characteristic spectra of Munsell colors. *Journal of the Optical Society of America, 6(2)*, 318-322.

Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. (2014). Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. *PLoS ONE 9(2:e87989)*, e87989.

Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding and Efficiency* (pp. 3-24).

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A, 16(3)*, 647-653.

Prins, N., & Kingdom, F. A. A. (2018). Applying the model-comparison approach to test specific tesearch hypotheses in psychophysical research using the Palamedes toolbox. *Frontiers in Psychology, 9*, 1250.

Radonjic, A., & Brainard, D. H. (2016). The nature of instructional effects in color constancy. *J Exp Psychol Hum Percept Perform, 42(6)*, 847-865.

Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015). Color constancy supports cross-illumination color selection. *Journal of Vision, 15(6)*, 1-19.

Radonjić, A., Ding, X., Krieger, A., Aston, S., Hurlbert, A. C., & Brainard, D. H. (2018). Illumination discrimination in the absence of a fixed surface-reflectance layout. *Journal of Vision, 18(5:11)*

Radonjić, A., & Gilchrist, A. L. (2013). Depth effect on lightness revisited: the role of articulation, proximity and fields of illumination. *i-Perception, 4(6)*, 437–455.

Radonjić, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N. P., Brainard, D. H., & Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal of Vision, 16(11:2)*, 1-18.

Rodieck, R. W. (1998). *The First Steps in Seeing*. Sunderland, Mass.: Sinauer.

Rovamo, J., Franssila, R., & Nasanen, R. (1992). Contrast sensitivity as a function of spatial frequency, viewing distance and eccentricity with and without spatial noise. *Vision Research, 32(4)*, 631-637.

Rovamo, J., Raninen, A., & Donner, K. (1999). The effects of temporal noise and retinal luminance on foveal flicker sensitivity. *Vision Research, 39*, 533-539.

Ruff, D. A., Ni, A. M., & Cohen, M. R. (2018). Cognition as a window into neuronal population space. *Annual Review of Neuroscience, 41*, 77-97.

Sankeralli, M. J., & Mullen, K. T. (1997). Postreceptoral chromatic detection mechanisms revealed by noise masking in three-dimensional cone contrast space. *Journal of the Optical Society of America A, 14(10)*, 2633-2646.

Schultz, S., Doerschner, K., & Maloney, L. T. (2006). Color constancy and hue scaling. *J Vis, 6(10)*, 1102-1116.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience, 16*, 1486-1510.

Shen, L., & Yeo, C. (2011). *Intrinsic images decomposition using a local and global sparse representation of reflectance*. Presented at the IEEE Conference on Computer Vision and Pattern Recognition,

Singh, V., Cottaris, N. P., Heasly, B. S., Brainard, D. H., & Burge, J. (2018). Computational luminance constancy from naturalistic images. *Journal of Vision, 18(13)*, 19.

Smithson, H. E. (2005). Sensory, computational, and cognitive components of human color constancy. *Philosophical Transactions of the Royal Society of London. Series B, 360(1458)*, 1329-1346.

Teller, D. Y. (1984). Linking propositions. *Vision Research, 24(10)*, 1233-1246.

Thibos, L. N., Hong, X., Bradley, A., & Cheng, X. (2002). Statistical variation of aberration structure and image quality in a normal population of healthy eyes. *Journal of the Optical Society of America A, 19(12)*, 2329-2348.

von Kries, J. (1905). Influence of adaptation on the effects produced by luminous stimuli. In D. L. MacAdam (Ed.), *Sources of Color Science (1970)* (pp. 120-1126). Cambridge, MA: MIT Press.

Vrhel, M. J., Gershon, R., & Iwan, L. S. (1994). Measurement and analysis of object reflectance spectra. *Color Research & Application, 19(1)*, 4-9.

Wandell, B. A., & Brainard, D. H. (in press). Principles and consequences of the initial visual encoding. In F. G. Ashby, H. Colonius & E. Dzhafarov (Eds.), *The New Handbook of Mathematical Psychology* Cambridge: Cambridge University Press.

Webster, M. A., & Mollon, J. D. (1995). Colour constancy influenced by contrast adaptation. *Nature, 373(6516)*, 694-698.

Weiss, D., Witzel, C., & Gegenfurtner, K. (2017). Determinants of colour constancy and the blue bias. *i-Perception, 8(6)*, 204166951773963.

Whittle, P., & Challands, P. D. C. (1969). The effect of background luminance on the brightness of flashes. *Vision Research, 9(9)*, 1095-1110.

Witzel, C., & Gegenfurtner, K. R. (2018). Color perception: objects, constancy, and categories. *Annual Review of Vision Science, 4*, 475-499.

Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: tests of three candidate cues. *Vision Research, 41*, 2581-2600.

Zhang, X., & Brainard, D. H. (2004). *Bayesian color correction method for non-colorimetric digital image sensors.* Paper presented at Color and Imaging Conference, 308-314.

Zhu, H., Yuille, A., & Kersten, D. (2021). *Three-dimensional pose discrimination in natural images of humans*. Presented at the Annual Meeting of the Vision Sciences Society, May 21-26, 2021. Poster A70.

1. We adopt the lightness discrimination threshold terminology based on the underlying assumption that observers perform the task using their perceptual lightness representation, and indeed our instructions to subjects used the lightness terminology to describe what should be judged. The actual stimulus variable being varied, however, was the simulated achromatic reflectance of the target object being judged, and feedback was given based on the value of this reflectance. In this paper, we do not explore the question as to whether the results would be affected if we had varied the instructions given to subjects (see footnote 1 above). [↑](#footnote-ref-1)
2. We use LRF rather than the more generic term albedo as our single number summary of the underlying spectral surface reflectance function, as the LRF is explicit about how variation in reflectance over wavelength should be taken into account. [↑](#footnote-ref-2)
3. The preregistration documents relevant to this paper are those for Experiments 6, 7 and 8. The site also contains preregistrations for previously reported (Experiment 1, 2 and 3, [cite equivalent noise paper]) and unreported (Experiment 4 and 5) work. [↑](#footnote-ref-3)
4. Here we neglect the effect of the fact that we truncated the distribution to enforce a requirement that reflectance at each wavelength lies between 0 and 1. We return to account for this in the LINRF formulation below. [↑](#footnote-ref-4)