

1 Title: Equivalent noise characterization of human lightness constancy  
2 Authors: Vijay Singh<sup>1,2</sup>, Johannes Burge<sup>2,3,4,5</sup>, David H. Brainard<sup>2,3</sup>  
3 1 Department of Physics, North Carolina A&T State University, Greensboro, NC, USA.  
4 2 Computational Neuroscience Initiative, University of Pennsylvania, Philadelphia, PA, USA.  
5 3 Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA.  
6 4 Neuroscience Graduate Group, University of Pennsylvania, Philadelphia, PA, USA.  
7 5 Bioengineering Graduate Group, University of Pennsylvania, Philadelphia, PA, USA.  
8  
9

10 **ABSTRACT:** A goal of visual perception is to provide stable representations of task-relevant scene  
11 properties (e.g. object reflectance) despite variation in task-irrelevant scene properties (e.g. illumination,  
12 reflectance of other nearby objects). To study such representational stability in the context of lightness  
13 representations in humans, we introduce a threshold-based psychophysical paradigm. We measure how  
14 thresholds for discriminating the achromatic reflectance of a target object (task-relevant property) in  
15 naturalistic scenes rendered using computer graphics are impacted by variation in the reflectance  
16 functions of background objects (task-irrelevant property). We refer to these thresholds as lightness  
17 discrimination thresholds. Our approach has roots in the equivalent noise paradigm. This paradigm relates  
18 signals to internal and external sources of noise and has been traditionally used to investigate contrast  
19 coding. We observe that, for low variation in background reflectance, lightness discrimination thresholds  
20 were nearly constant, indicating that observers' internal noise determines threshold in this regime. As  
21 background object reflectance variation increases, its effects start to dominate performance. We report  
22 lightness discrimination thresholds as a function of the amount of variability in the background object  
23 reflectance function to determine the equivalent noise - the smallest level of task-irrelevant (i.e.  
24 background reflectance) variation that substantially corrupts the visual representation (i.e. perceived  
25 object lightness) of the task-relevant variable (i.e. achromatic reflectance). A linear receptive field model,  
26 which employs a single center-surround receptive field tailored to our stimulus set, captures human  
27 behavior in this task. Our approach provides a method for characterizing the effect of task-irrelevant  
28 scene variations on the perceptual representation of a task-relevant scene property.

29 **KEYWORDS:** Lightness, Equivalent Noise, Human Psychophysics, Color Vision  
30

1    **1 INTRODUCTION**

2  
3    To support effective action, vision provides stable perceptual representations of the distal properties of  
4    objects. The computations that give rise to these representations start with the information in the proximal  
5    stimuli reaching the retinas. These proximal stimuli depend on the observer's particular viewpoint on the  
6    scene, on object-extrinsic properties of the scene (e.g. illumination), and on the intrinsic properties of the  
7    objects in the scene. The challenge for the visual system is to recover stable correlates of object-intrinsic  
8    properties across variation in other scene variables. Understanding the degree to which the visual system  
9    rises to this challenge, and how it does so, is an important goal of vision science (Helmholtz, 1896; Knill  
10   & Richards, 1996; Brascamp & Shevell, 2021).

11  
12   Here we consider the perceptual task of representing the reflectance of an object embedded in a scene  
13   based on the light reflected to the eye from the object as well as the rest of the scene. The perceptual  
14   correlate of the surface reflectance of an object is its perceived color or, in the special case of achromatic  
15   objects, its lightness. Computing a stable color and lightness representation poses a challenge to the visual  
16   system because the retinal image of the object varies with the object's reflectance, the spectral irradiance  
17   of the illumination, and the position and pose of the object in the scene. The degree that the visual system  
18   succeeds at stabilizing its object color and lightness representation in the face of this variation determines  
19   the degree to which we say that the visual system achieves color and lightness constancy.

20  
21   Color and lightness constancy have been studied extensively in experiments where observers report the  
22   perceived color or lightness of a target object, across changes extrinsic to the target object reflectance.<sup>1</sup>  
23   Here, the target object's reflectance is the task-relevant scene variable, while other aspects of the scene  
24   are task irrelevant. Under some viewing conditions, the visual system can achieve high degrees of  
25   constancy in the face of changes in illumination (Foster, 2011). This constancy is mediated by several  
26   different cues, including the mean and variance of the light reflected from the overall scene, from surfaces  
27   neighboring the target surface, and from the more luminous regions of the scene (Smithson, 2005;  
28   Brainard & Radonjić, 2014; Witzel & Gegenfurtner, 2018; Hurlbert, 2019). Several theoretical  
29   frameworks have been developed, which provide a variety of means for understanding how different cues  
30   shape perceptual representations of object reflectance (see reviews cited earlier in this paragraph as well  
31   as Adelson, 2000; Gilchrist, 2006; Kingdom, 2011; Brainard & Maloney, 2011; Murray, 2021).

32  
33   Psychophysical methods for measuring discrimination thresholds complement the class of experiments  
34   described above. Thresholds do not provide direct measurements of perceived color or lightness, but they  
35   can be used to assess the precision of the underlying perceptual representation. Psychophysical threshold  
36   measurements are accompanied by mature theory that can be used to link the measurements to properties  
37   of physiologically measured neural responses (Brindley, 1960; Green, 1996; Teller, 1984; Parker &  
38   Newsome, 1998). In addition, threshold measurements can be readily adapted for use with non-human  
39   subjects, since trial-by-trial reward can be provided based on whether each response is correct. It is less  
40   clear, however, how to apply threshold measurements to questions of perceptual constancy. One approach  
41   is to connect thresholds to suprathreshold judgments of perceived stimulus properties, an approach which  
42   has its origins in Fechner's pioneering interpretation of Weber's Law (Fechner, 1966). The basic idea is  
43   that both threshold and suprathreshold percepts are mediated by a common stimulus-response function  
44   whose properties depend on and can change with viewing context. Thresholds are related to the slope of  
45   the response function, with higher slopes leading to larger response changes for a fixed stimulus change

---

<sup>1</sup> This type of experiment may be instrumented with instructions that prompt the observer to report how the object appears, or with instructions that prompt the subject to report their estimate of the object's reflectance. Exactly what observers report under either of these instructional regimes, as well as the nature of instructional effects, is an important but thorny issue that we will not digress on further in this paper. See Radonjić and Brainard (2016) for a recent treatment of the issue, as well as the references therein.

1 and thus lower thresholds. The magnitude of a suprathreshold percept, on the other hand, is related to the  
2 value of the response function produced by the stimulus being judged. Although positing a common  
3 stimulus-response function holds promise (Nachmias & Sansbury, 1974; Hillis & Brainard, 2005; Hillis  
4 & Brainard, 2007b), there are documented cases where the threshold measurements are inconsistent with  
5 measures of lightness constancy (Hillis & Brainard, 2007a).

6  
7 Another threshold-based approach to constancy is to study the ability to detect a change in the  
8 confounding scene property (e.g., the illumination; Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert,  
9 2014; Radonjić et al., 2016; Radonjić et al., 2018; Aston, Radonjić, Brainard, & Hurlbert, 2019; Alvaro,  
10 Linhares, Moreira, Lillo, & Nascimento, 2017), rather than a change in the object property of interest (e.g.  
11 surface reflectance). The connection to constancy is via the linking proposition that for sub-threshold  
12 illumination changes, the visual system cannot detect any change in the image and thus that the perceptual  
13 representation of surface reflectance, as well as of illumination, is stable across such sub-threshold  
14 illumination changes. However, how the results of illumination discrimination thresholds, which probe  
15 the effect of small illumination changes, relate to the stability of perceptual judgments across larger  
16 illumination changes that occur in natural viewing has not been worked out (but see Weiss, Witzel, &  
17 Gegenfurtner, 2017).

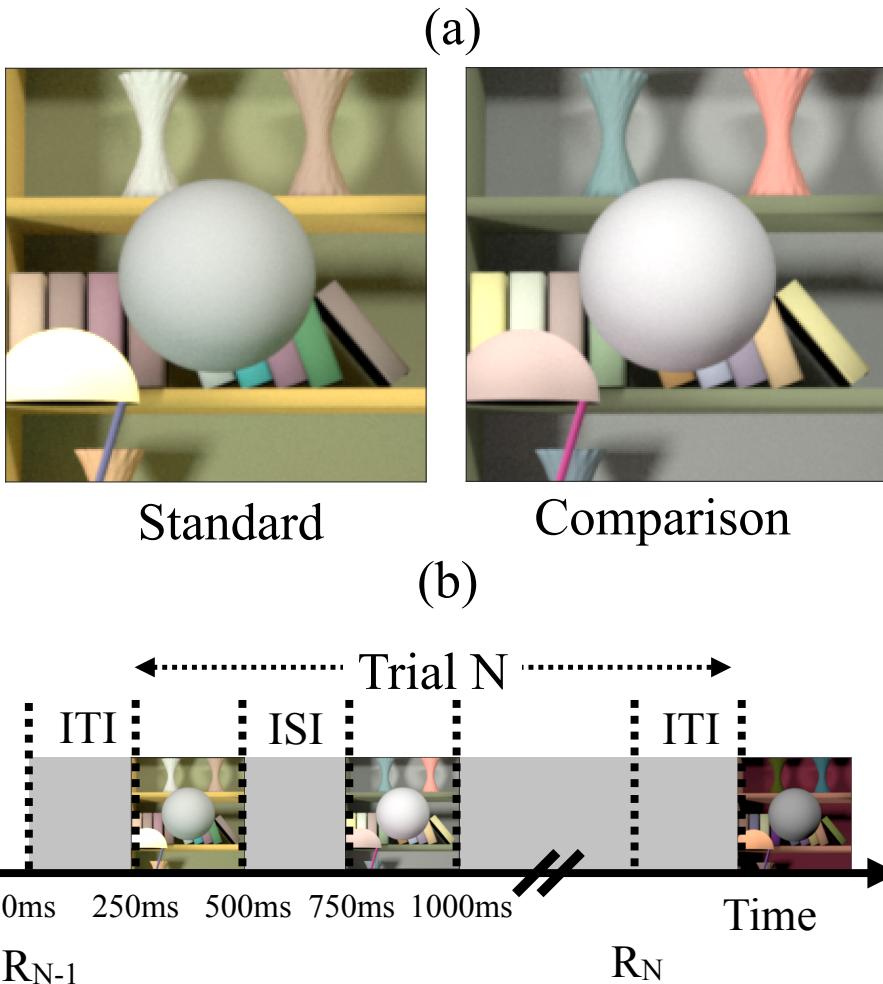
18  
19 Here we introduce a new approach to using a psychophysical threshold paradigm to draw inferences  
20 about perceptual constancy and its underlying mechanisms, based on measuring how thresholds for a  
21 task-relevant scene property are affected by variation in a task-irrelevant scene property. This approach is  
22 conceptually similar to studying how thresholds are affected by addition of spatially white or pink noise  
23 (Legge, Kersten, & Burgess, 1987; Pelli, 1990; Pelli & Farell, 1999), but with the noise introduced as  
24 variation in distal scene properties. In this paper, we apply this approach towards understanding lightness  
25 constancy in naturalistic scenes rendered using computer graphics, but the ideas are general. First, we  
26 measure human ability to discriminate the achromatic surface reflectance of two objects in the absence of  
27 any object-extrinsic variation. Next, we measure how these discrimination thresholds, which we refer to  
28 as lightness discrimination thresholds, change with the introduction of object-extrinsic variation in the  
29 form of variability in the reflectance spectra (loosely, the colors) of the background objects in the scene  
30 (Lotto & Purves, 1999; Brown & MacLeod, 1997). The discrimination threshold at each level of  
31 background variation quantifies the difficulty of the lightness discrimination task in each condition. The  
32 change in thresholds from baseline (i.e., no background variation) quantifies the degree to which the  
33 object-extrinsic variation intrudes on the object-intrinsic representation. More generally, we can think of  
34 this as a paradigm in which thresholds for discriminating a task-relevant stimulus variable, here  
35 achromatic surface reflectance, are measured in the face of variation of a task-irrelevant stimulus variable,  
36 here the reflectance of the background objects.

37  
38 As the variation in background surface reflectance is increased, discrimination thresholds are constant and  
39 then increase, with log squared threshold increasing linearly with log background surface reflectance  
40 variance. The minimum discrimination threshold and the background surface reflectance variance at  
41 which the threshold begins to rise are consistent across different observers. Moreover, a simple model  
42 rooted in Signal Detection Theory, that makes use of a single center-surround receptive field captures the  
43 essential features of the psychophysical data. This approach allows us to quantify the effect of extrinsic  
44 variation on the observer's representation of lightness, relative to the intrinsic precision of that variation.

## 45 2 RESULTS

### 46 2.1 Measurement of lightness discrimination thresholds

47  
48 We measured how variation in the reflectance spectra of background objects affects thresholds for  
49 discriminating object achromatic reflectance, which we refer to as lightness discrimination thresholds,



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

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

1 using a two-alternative forced-choice (2AFC) design (Figure 1). On each trial, observers viewed a  
2 standard image and comparison image, sequentially presented on a calibrated monitor for 250ms each,  
3 with a 250ms inter-stimulus interval (Figure 1a). The images were computer graphics renderings of 3D  
4 scenes. Each scene contained an achromatic spherical target object. The task was to report the image in  
5 which the depicted target object was lighter. Across trials, we varied the luminous reflectance factor  
6 (LRF; American Society for Testing and Materials, 2017) of the target object in the comparison image  
7 while keeping the LRF of the target object in the standard image fixed. The LRF is the ratio of the  
8 luminance of a surface under a reference illuminant (here, we used the CIE D65 reference illuminant) to  
9 the luminance of the reference illuminant itself. Feedback was given on each trial based on which image  
10 contained the target object with the higher LRF.<sup>2</sup>

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

17

## 18 **2.2 Human lightness discrimination thresholds increase with background object reflectance 19 variation**

20 To study the effect of background variation on lightness discrimination thresholds, we varied the  
21 reflectance spectra of the background objects in the images by sampling from a statistical model built  
22 from natural surface reflectance databases (see Methods: Reflectance and Illumination Spectra Singh,  
23 Cottaris, Heasly, Brainard, & Burge, 2018). Briefly, a database of natural surface reflectance functions  
24 (Kelly, Gibson, & Nickerson, 1943; Vrhel, Gershon, & Iwan, 1994) was projected along eigenvectors  
25 associated with the largest six eigenvalues of the dataset. These eigenvalues captured more than 90% of  
26 the variance in the database. The distribution of projection weights was approximated as a multivariate-  
27 normal distribution. Reflectance spectra of background objects were constructed by first sampling from  
28 the multivariate distribution of projection weights and then performing a weighted sum of the  
29 corresponding eigenvectors. The amount of variation in the background was controlled by multiplying the  
30 covariance matrix of the multivariate-normal distribution by a scalar.

31 We measured threshold as a function of the scalar that multiplied the covariance matrix. By varying the  
32 scalar from 0 (no variation) to 1 (variation in natural scenes), we examined parametrically how  
33 background variation affects performance in the task. We generated images for six logarithmically spaced  
34 values of the covariance scalar. Figure 3 shows examples of images used in our psychophysical task for  
35 different choices of the covariance scalar. Discrimination thresholds were measured separately for each of  
36 the six values of the covariance scalar (Appendix: Table S2).

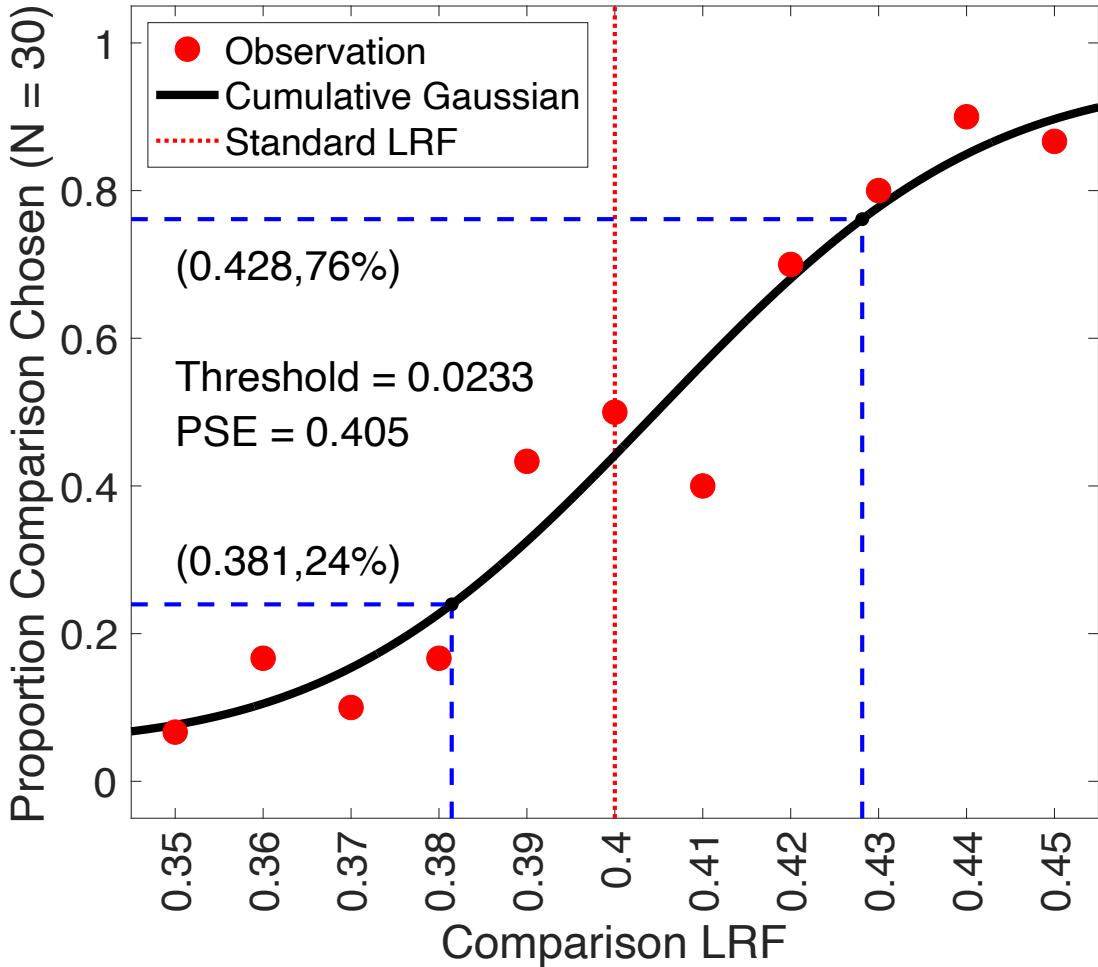
37

38 Figure 4 shows how discrimination thresholds change with the amount of variability in the spectra of the  
39 background objects. Mean log threshold squared (averaged across observers, N = 4) is plotted against the  
40 log of the covariance scalar. For low values of the covariance scalar, the thresholds are nearly constant.  
41 As the covariance scalar increases, log squared threshold rises approximately linearly with log covariance  
42

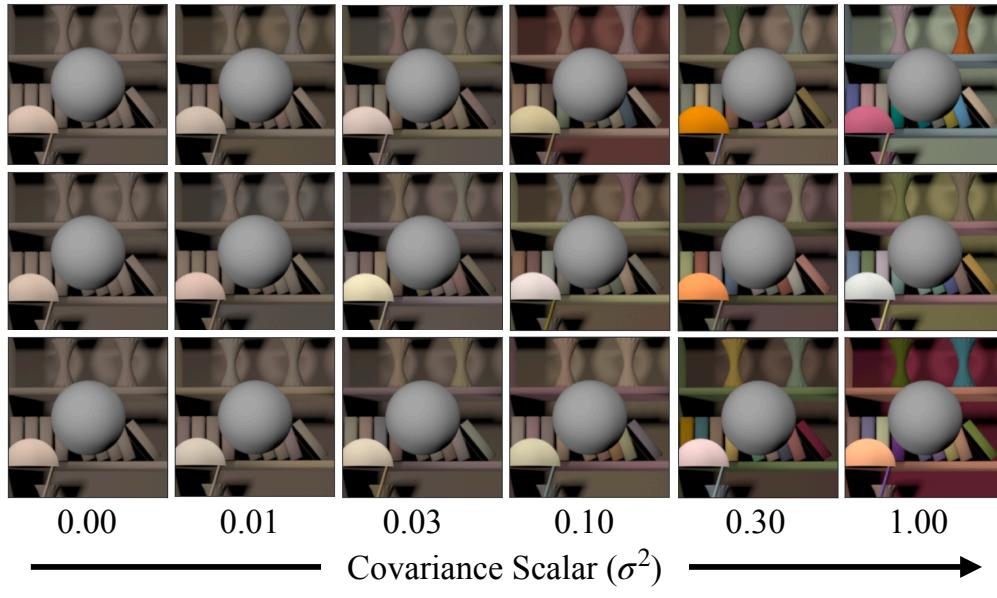
43

---

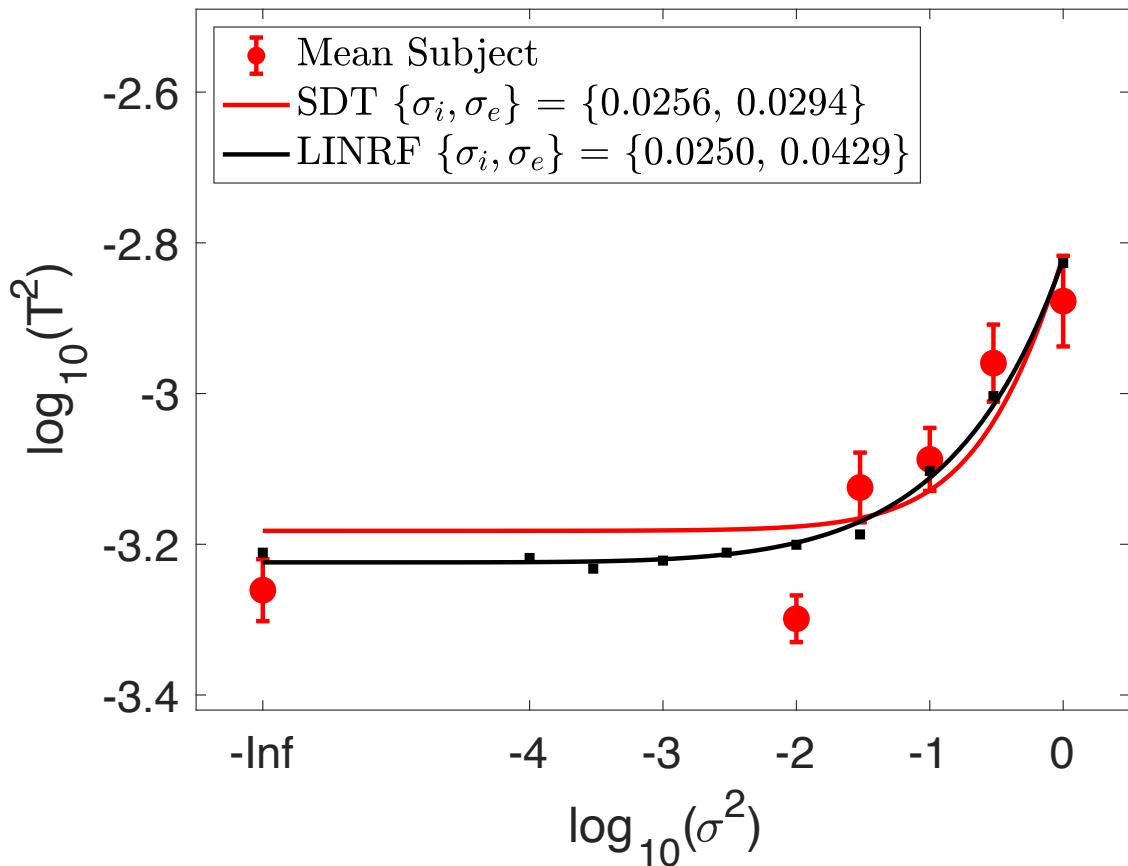
44 <sup>2</sup> We adopt the lightness discrimination threshold terminology based on the underlying assumption that observers  
45 perform the task using their perceptual lightness representation, and indeed our instructions to subjects used the  
46 lightness terminology to describe what should be judged. The actual stimulus variable being varied, however, was  
47 the simulated achromatic reflectance of the target object being judged, and feedback was given based on the value of  
48 this reflectance. In this paper, we do not explore the question as to whether the results would be affected if we had  
49 varied the instructions given to subjects (see footnote 1 above).



**Figure 2: Psychometric function.** We recorded the proportion of times the observer chose the target in the comparison image to be lighter, as a function of the comparison LRF. The LRF of the target object in the standard image was fixed at 0.4. The LRF of the target object in the comparison image were chosen from 11 linearly spaced values in the range [0.35, 0.45]. Thirty trials were presented at each comparison LRF value. We fit a cumulative normal distribution to the proportion comparison chosen data using maximum likelihood methods. The guess and lapse rates were assumed to be equal and were restricted to be in the range [0, 0.05]. The threshold was measured as the difference between the LRF at proportion comparison chosen equal to 0.7604 and 0.5, as predicted by the cumulative normal fit. This figure shows the data for Observer 2 for scale factor 0.00 in the first experimental session for that observer. The point of subjective equality (PSE, the LRF corresponding to proportion chosen 0.5) was close to 0.4 as expected and the threshold was 0.0233. The lapse rate for this fit was 0.05.



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



**Figure 4: 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  $\pm 1$  SEM taken between observers. The data were fit with the function (SDT Model)  $T^2 = T_0^2 (\sigma_i^2 + \sigma^2 \sigma_e^2)$  with  $T_0 = 1$  (red curve). The best fit parameters are indicated in the legend. The threshold of the linear receptive field (LINRF) model was estimated at 10 logarithmically spaced values of the covariance scalar (black squares). The black smooth curve is a smooth fit to these points of the functional form  $\log_{10} T^2 = a + b^{(x+c)^d}$  where  $x = \log_{10} \sigma^2$  and  $a, b, c$  and  $d$  are parameters adjusted in the fit.

scalar, a dependence predicted by a simple model based on Signal Detection Theory (Figure 4; see below and Methods: Signal Detection Theory Model).

When the covariance scalar is 0, we conceptualize performance as limited by two factors (Pelli & Farell, 1999). One factor is the internal variability in the observer's representation of target object lightness. The other factor is the efficiency with which the observer's decision processes make use of the information provided by this representation. Our experiments cannot identify the relative contributions of these two conceptually distinct factors. In the following, we refer to both factors collectively as the observer's internal noise for the lightness discrimination task.

As the covariance scalar increases, a third factor becomes important. This factor is the impact of background object surface variability on the observer's representation of target object lightness. At low values of the covariance scalar, internal noise dominates, and external variability has little effect on threshold. At high values of the covariance scalar, external variability limits performance, and thresholds increase systematically with increase in the covariance scalar. We interpret these effects further in the context of modeling introduced below.

Figure 5 shows threshold variation for each individual observer. The pattern of thresholds for each individual observer follows the same basic pattern as the mean thresholds across observers. Thresholds are constant for low values of the covariance scalar, and, for higher values of the covariance scalar, thresholds rise approximately linearly on the log threshold squared versus log covariance plot. The most notable difference across individual observers is the slope of the rising limb of the measured functions.

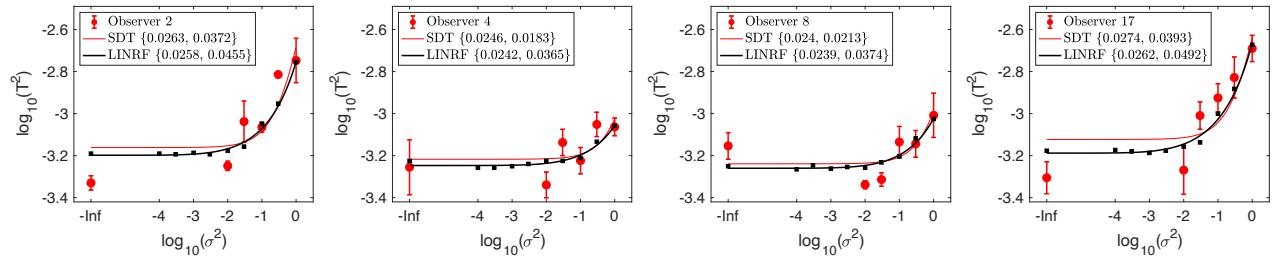
### 2.3 Quantifying impact of background surface variation on the lightness representation

We modeled the psychophysical data with a framework based on Signal Detection Theory (SDT, see Methods: Signal Detection Theory Model). Our framework posits that performance is limited by two fundamental factors. The first factor is variability internal to the visual system (i.e., internal noise). The second factor is the effect of experimentally induced stimulus variability (i.e., background surface-reflectance variability) on the visual system's representation of lightness (external noise). The models aid in evaluating how much external noise intrudes on performance, compared to the intrinsic precision of the visual system's representation of target lightness. The model relates the discrimination threshold ( $T$ ) with the variance in the internal noise ( $\sigma_i^2$ ), the external noise ( $\sigma_e^2$ ), and the covariance scalar ( $\sigma^2$ ) as:

$$T^2 = T_0^2 (\sigma_i^2 + \sigma^2 \sigma_e^2) \quad (1)$$

where  $T_0$  is the threshold with no external variation (see Methods: Signal Detection Theory Model for details). Performance with no external variation (covariance scalar,  $\sigma^2 = 0$ ) establishes the level of the internal noise, while the covariance scalar value corresponding to the threshold that is twice the threshold with no external variation indicates when the amount of external noise matches the amount of internal noise.

To relate the SDT model directly to the stimuli used in our experiments, we developed a version of the SDT model based on a single-channel linear receptive field (see Methods: Linear Receptive Field Model; LINRF). This version of the model calculates the response of a linear receptive field to an image to instantiate the internal representation for the LRF postulated in the SDT model. This representation is then used in a simulation of the 2AFC paradigm to estimate model threshold. The receptive field model provides a straightforward way to incorporate features of human physiology. Here, these include optical blur (Westheimer, 1986) and the Poisson noise that perturbs cone photoreceptor isomerizations in the retina (Rodieck, 1998). In addition, this model allows us to easily account for the details of the rendering



**Figure 5: Threshold of individual human observers.** Mean (across sessions) squared threshold vs log covariance scalar for individual human observers. Same format as Figure 4; here the error bars represent  $\pm 1$  SEM taken across sessions for each observer. The parameters of the SDT model and the linear receptive field (LINRF) models were obtained separately for each observer.

6

1 process used to generate the images for the experiment, such as the truncation of surface reflectances to  
2 the range 0 to 1 in our model of natural surface reflectances.

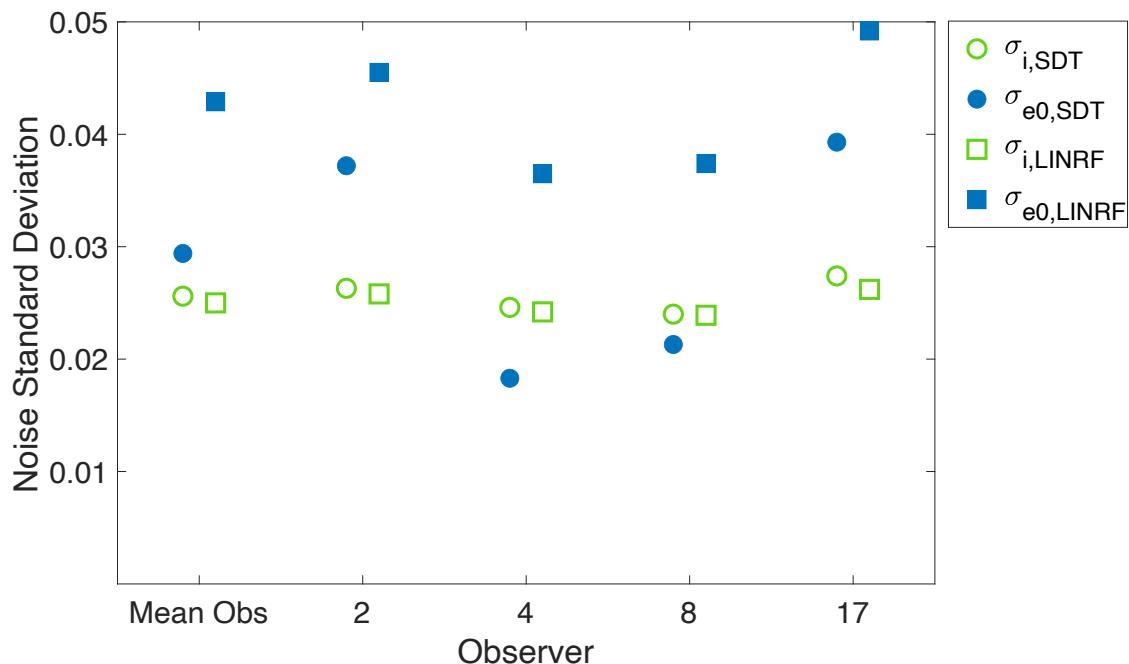
3  
4 Figure 4 shows the fit of the SDT model and the linear receptive field model to the mean observer data.  
5 Figure 5 shows the model fits to the individual observer data. Both versions of the model capture the  
6 broad features of the data. The linear receptive field model provides a better fit, because this model  
7 enables us to account for the fact that the actual covariance of the variation in background surface  
8 reflectances differs from the nominally specified variation, because we enforce a physical realizability  
9 constraint that surface reflectances lie between 0 and 1 (See Methods: Reflectance and Illumination  
10 Spectra).

11  
12 The model fits provide estimates of internal and external noise for the human observers in this task.  
13 Figure 6 plots the estimates of the internal and external noise standard deviations (quantities  $\sigma_i$  and  $\sigma_{e0}$ ,  
14 see Methods) for the SDT model and the linear receptive field model.

15  
16 The estimates of the internal noise are similar for the two models, which makes sense since for zero  
17 external noise the model formulations converge; the Poisson noise included in the linear receptive field  
18 model does not typically limit human discrimination performance at daylight light levels (Banks, Geisler,  
19 & Bennett, 1987; Cottaris, Jiang, Ding, Wandell, & Brainard, 2019). The mean values of the internal  
20 noise standard deviation are close to the values obtained by fitting the mean data (SDT model: mean  
21 value of internal noise standard deviation of individual observers 0.0256, value from fit to mean data  
22 0.0256; linear receptive field model: mean value of internal noise standard deviation of individual  
23 observers 0.0250, value from fit to mean data, 0.0250).

24  
25 The estimates of external noise are higher for the linear receptive field model than for the SDT model  
26 (SDT model: mean value of external noise standard deviation 0.0290, value from fit to mean data 0.0294;  
27 linear receptive field model: mean value of external noise standard deviation 0.0421, value from fit to  
28 mean data, 0.0429). This is consistent with the observation that the SDT model underestimates the rise in  
29 thresholds with increasing covariance scalar, while this rise is captured accurately by the linear receptive  
30 field model, presumably because the latter incorporates the constraint that the surface reflectance values  
31 at each wavelength are physically realizable (i.e., surface reflectances lie between 0 and 1). If we focus on  
32 the estimates from the linear receptive field model fit to the mean data,  $\sigma_{e0}$  is larger by a factor of  $\sim 1.7$   
33 than  $\sigma_i$ . To the extent that our model of natural surface reflectance is accurate, this tells us that the  
34 variance of human representations of lightness, given substantial variation in background surface  
35 reflectance, are within a factor of two of the limits imposed by the intrinsic precision of that  
36 representation.

37



1  
2 **Figure 6: Internal and external noise standard deviation for human observers.** Noise standard  
3 deviation for human observers estimated using SDT model and the computational linear receptive model  
4 (LINRF) model. While the internal noise estimates are consistent over the two models, the external noise  
5 estimated by the LINRF model is higher compared to the SDT model.

6

1      **3 DISCUSSION**

2  
3      The perceived lightness of an object can depend on the scene in which it lies. Stabilization of the lightness  
4      representation against variation in scene properties extrinsic to the object's surface reflectance is referred  
5      to as lightness constancy. In this paper, we introduced a psychophysical approach to characterize lightness  
6      constancy based on measuring how lightness discrimination thresholds vary with experimentally  
7      introduced variation in scene properties extrinsic to the object's reflectance. Specifically, we studied how  
8      lightness discrimination thresholds are impacted by variation in the reflectance of the background objects  
9      in naturalistic scenes rendered using computer graphics. Our results (Figures 4 and 5) show that when the  
10     variation in the surface reflectance of background objects is small, discrimination thresholds are nearly  
11     constant. In this regime, performance is limited primarily by internal noise. As the amount of background  
12     reflectance variation increases, the effect of external variation starts dominating that of the internal noise,  
13     and discrimination thresholds increase. We analyzed the data using a modeling approach used previously  
14     to study effect of external noise on contrast detection (Legge, Kersten, & Burgess, 1987; Pelli, 1990; Pelli  
15     & Farell, 1999). This approach allows us to relate the effect of background surface variation to the  
16     intrinsic precision of the lightness representation. We find that the effect of the external variability  
17     introduced by variation of background surface reflectances in naturalistic scenes is within a factor of two  
18     of the intrinsic precision of the lightness representation. More generally, our work provides a method to  
19     quantify the effect of variation in task-irrelevant properties on the perception of task-relevant property.  
20

21     *Spatial and chromatic properties of the stimuli.* We used small image patches in our study, an important  
22     difference between our stimuli and natural viewing. In this initial deployment of our paradigm, we  
23     focused on effects of background variation that are relatively nearby the test object, and which are likely  
24     mediated by relatively small populations of neurons. The use of small image patches is not a necessary  
25     requirement of our paradigm. Extending the work to larger patches is an obvious next step. In addition to  
26     using small patches, we did not vary the spatial structure of the array of objects in the rendered scenes.  
27     Manipulating spatial structure may provide a way to use our paradigm to measure the spatial tuning of the  
28     mechanism(s) mediating the background effect. This approach is loosely analogous to how manipulating  
29     the structure of contrast noise may be used to examine the tuning of mechanisms supporting the detection  
30     of contrast-defined targets (Henning, Hertz, & Hinton, 1981; Rovamo, Franssila, & Nasanen, 1992;  
31     Losada & Mullen, 1995; Nachmias, 1999; Rovamo, Raninen, & Donner, 1999). Similarly, it may be  
32     possible to manipulate the chromatic structure of the variation in background surface reflectances with the  
33     goal of understanding the chromatic tuning of the background effect. This would again be analogous to  
34     how noise-based approaches have been used to characterize chromatic tuning of mechanisms that support  
35     the detection of chromatically-defined contrast targets (Gegenfurtner & Kiper, 1992; Sankeralli &  
36     Mullen, 1997; Giulianini & Eskew, 1998; Monaci, Menegaz, Süsstrunk, & Knoblauch, 2004).

37  
38     *Link between thresholds and suprathreshold perceptual judgments.* The technique developed here probes  
39     the constancy of a perceptual representation of a task-relevant variable (e.g., perceived object lightness)  
40     by measuring how variation in a task-irrelevant scene variable (e.g., background surface reflectance)  
41     elevates thresholds for detecting changes in the task-relevant variable. As with other threshold-based  
42     methods for approaching the stability of suprathreshold perceptual judgments (see Introduction), the  
43     extent to which the results may be used to predict the stability such judgments across changes in other  
44     scene variables is not known. Experiments that explore this link are of considerable interest.  
45

46     *Applications to understanding neural mechanisms.* A longstanding goal of vision science is to connect  
47     psychophysical performance to its underlying neural mechanisms. For probing mechanisms that mediate  
48     perceptual constancies, our paradigm has the attractive feature that there is a well-defined correct answer  
49     on each trial, so that for studies with animal subjects it is possible to provide performance-contingent  
50     reward. In addition, there are well-worked out methods for predicting psychophysical discrimination  
51     performance from recordings of the responses of neural populations (Shadlen, Britten, Newsome, &

1 Movshon, 1996; Parker & Newsome, 1998; Cohen & Newsome, 2009; Nienborg, Cohen, & Cumming,  
2 2012; Ruff, Ni, & Cohen, 2018), and the theoretical links between such analysis and performance should  
3 continue to hold when task-irrelevant stimulus variation is added to the paradigm. Complementing these  
4 measurements with normative analyses that leverage receptive fields that encode the optimal stimulus  
5 features for specific tasks and that optimally decode the responses of these receptive fields into estimates  
6 (or forced-choice responses) will help enrich our understanding of the links between sensory-perceptual  
7 processing, neural computation, and psychophysical performance (Geisler, Najemnik, & Ing, 2009; Burge  
8 & Jaini, 2017; Jaini & Burge, 2017; Burge, 2020). There have already been some successes of this  
9 approach in the domains of blur, binocular disparity, and speed estimation in naturalistic images (Burge &  
10 Geisler, 2011; Burge & Geisler, 2014; Burge & Geisler, 2015; Chin & Burge, 2020). We are excited  
11 about the potential of our paradigm to provide rigorous quantitative insights about the sensory-perceptual  
12 processing and the neural computations underlying color constancy in particular, and perceptual  
13 constancy more generally. Our previous computational work on lightness constancy is relevant in this  
14 context (Singh, Cottaris, Heasly, Brainard, & Burge, 2018).

15  
16 *Model of natural surface reflectances.* We used a statistical model of naturally-occurring surface  
17 reflectances to determine the distribution from which we sampled the background surface reflectance  
18 functions. This model was developed in our earlier work (Singh, Cottaris, Heasly, Brainard, & Burge,  
19 2018; see also Brainard & Freeman, 1997; Zhang & Brainard, 2004). The model is based on  
20 measurements of surface reflectance functions of the Munsell papers (Kelly, Gibson, & Nickerson, 1943)  
21 as well as natural surfaces characterized by Vrhel (1994). Although we view this as a reasonable model, it  
22 is important to note that the quantitative relation we measured between the magnitude of internal noise  
23 and the effect of external noise for natural surface reflectances depends on how the distribution of natural  
24 surface reflectances is modeled. If the model of surface reflectance overestimates the variation in natural  
25 surfaces, the effect of external noise in natural scenes will be less than we estimated. Conversely, if the  
26 model of surface reflectance underestimates the variation in natural surfaces, the effect of external noise  
27 in natural scenes will be greater than we estimated (Singh, Cottaris, Heasly, Brainard, & Burge, 2018).  
28 Future refinement of surface reflectance models could be used in conjunction with the parameters of the  
29 linear receptive field model developed here, without need for new data collection, to refine the estimate of  
30 the effect of naturally occurring background variation on object lightness perception.  
31

32 *Rule of combination.* In the present work, we considered variation in only a single task-irrelevant variable.  
33 In natural scenes, there are many task-irrelevant variables. In the case of judging object lightness, these  
34 include object-extrinsic factors such as the scene illumination, the position and 3D orientation of the  
35 target object in the scene, the viewpoint from which the object is viewed, and various object-intrinsic  
36 factors like its shape and size. Variation in each of the factors could in principle elevate thresholds for  
37 discriminating object lightness. Our paradigm allows characterization of the effect of these task-irrelevant  
38 variables and quantifies that effect for each such variable in the same internal-noise referred units. One  
39 potentially important future direction is to measure the combined effect of simultaneous variation of  
40 multiple task-irrelevant variables, and to test hypotheses about the rules of combination.  
41

42  
43 **4 ACKNOWLEDGEMENTS:** NIH R01-EY10016 (DHB), NIH R01-EY028571 (JB).

44  
45 **5 METHODS**

46  
47 **5.1 Ethics statement**

48 All experimental procedures were approved by University of Pennsylvania Institutional Review Board  
49 and were in accordance with the World Medical Association Declaration of Helsinki.

1   **5.2 Preregistration**

2   The experimental design and the method for extracting threshold from the data were preregistered before  
3   the start of the experiment. They are publicly available at: <https://osf.io/7tgy8/>. Deviations from and  
4   additions to the preregistered plan are described in the addendums to the pre-registration documents  
5   available at <https://osf.io/7tgy8/>.

6   The broad aim of the study was to study the effect of object extrinsic scene variations on human object  
7   lightness discrimination thresholds. For this, we pre-registered three experiments. The first experiment  
8   (pre-registered as Experiment 1) was abandoned because the task was too difficult. The findings of the  
9   second experiment (pre-registered as Experiment 2) provided control data and are reported in the  
10   Appendix. We focus in the paper on the third experiment (pre-registered as Experiment 3).

11   A deviation from the pre-registered plan for pre-registered Experiment 2 was the change in the criteria to  
12   select observers for the experiment. The pre-registered criterion for selecting an observer for Experiment  
13   2 was that an observer would be excluded if their mean threshold for the last two acquisitions run in the  
14   practice session exceeded 0.025. After collecting data from 8 naive observers, we concluded that this  
15   criterion was too strict as only one observer met the criterion. Hence, we increased exclusion threshold  
16   from 0.025 to 0.030. The pre-registered plans also indicated that each image would be presented for  
17   500ms, but in the event we shortened this to 250ms.

18   We followed the procedure described in the pre-registration document to extract threshold from the data.  
19   The document also indicated that the primary data feature of interest was the dependence of threshold on  
20   the covariance scalar and predicted that thresholds would increase with increasing background variability.  
21   The quantitative models of the data, however, were developed post-hoc.

22   **5.2 Apparatus**

23   The stimuli were presented on a calibrated LCD color monitor (27-in. NEC MultiSync PA271W; NEC  
24   Display Solutions) in an otherwise dark room. The monitor was driven at a pixel resolution of 1920 x  
25   1080, a refresh rate of 60Hz, and with 8-bit resolution for each RGB channel. The host computer was an  
26   Apple Macintosh with an Intel Core i7 processor. The experimental programs were written in MATLAB  
27   (MathWorks; Natick, MA) and relied on routines from the Psychophysics Toolbox  
28   (<http://psychtoolbox.org>) and mgl (<http://justingardner.net/doku.php/mgl/overview>). Responses were  
29   collected using a Logitech F310 gamepad controller.

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

33   **5.3 Monitor Calibration**

34   The monitor was calibrated using a spectroradiometer (PhotoResearch PR650). To calibrate the monitor,  
35   we focused the spectroradiometer on a patch displayed on the center of the monitor. The patch size was  
36   4.66cm x 4.66cm (3.56° x 3.56°). The optics of the radiometer sampled the emitted light from a 1°  
37   circular spot within the patch. The spectral power distribution of the three monitor primaries was  
38   measured in the range 380nm to 780nm at 4nm steps. The gamma functions for each primary were  
39   determined from measurements of the spectral power distribution for each primary at 26 equally spaced  
40   input values for that primary, in the range [0, 1] where 1 corresponds to the maximum value of the  
41   allowed input and 0 corresponds to no input. These gamma functions as well as the light emitted by the

1 monitor for an input of 0 were accounted for in the stimulus display procedures (Brainard, Pelli, &  
2 Robson, 2002). The spectral power distribution of the three primaries were also measured at 32 different  
3 combinations of the input in the range [0,0,0] to [1,1,1]. These measurements were used check the  
4 linearity of the display. The maximum absolute deviation of the x-y chromaticity between the measured  
5 values and those predicted based on linearity was 0.0028 and 0.0027 for x and y chromaticity  
6 respectively, and less than 1% for luminance.

#### 7 **5.4 Observer Recruitment and Exclusion**

8 Observers were recruited from the University of Pennsylvania and the local Philadelphia community and  
9 were compensated for their time. Observers were screened to have normal visual acuity (20/40 or better)  
10 and normal color vision, as assessed with pseudo-isochromatic plates (Ishihara, 1977). These exclusion  
11 criteria were specified in the pre-registration document. One observer was discontinued at this point as  
12 they did not meet the normal visual acuity criterion.  
13

14 Observers who passed the vision screening then participated in a practice session. This session also served  
15 to screen for observers' ability to reliably perform the psychophysical task. This screening was performed  
16 in the first session for each observer, which was considered a practice session. At the beginning of the  
17 practice session, observers were familiarized with the task. For this they performed a familiarization  
18 acquisition (See Methods: Experimental Details for the definition of an acquisition). In the familiarization  
19 acquisition, observers performed 40 trials of the task using images with covariance scale factor 0.00 (10  
20 easy trials, 10 moderate trials, and 20 regular trials). In the easy trials, the observers compared images  
21 with target object luminous reflectance factor (LRF) 0.35 and 0.45. In the moderate trials, they compared  
22 images with target object LRF 0.40 to images with target object LRF 0.35 or 0.45. In the regular trials  
23 they compared images with target object LRF 0.40 to images with target object LRF in the range [0.35,  
24 0.45]. The data from the familiarization acquisition was not saved. After this the observer performed three  
25 normal acquisitions for images with covariance scale factor 0.00. At the end of the practice session, the  
26 mean threshold of the observer for the last two acquisitions was computed. The observer was excluded  
27 from further participation if their mean threshold for the last two acquisitions in the practice session  
28 exceeded 0.025 ( $\log T^2$ , -3.2). This exclusion criterion was specified in our pre-registered protocol.  
29

30 Observers who met the performance criterion participated in the rest of the experiment. Observers  
31 performed only one session on a given day. The sessions were scheduled as per the availability of the  
32 experimenter and the observer. The data of all observers in the main experiment (pre-registered  
33 Experiment 3) was collected over a period of 4 weeks.  
34

35 A total of 17 observers participated in the practice sessions for Experiments 2 and 3. To de-identify  
36 observer information in the data, observers were numbered in the order they performed the practice  
37 sessions. 10 observers participated in the practice sessions for Experiment 3 (6 Female, 4 Male; age 18-  
38 56; mean age 30.7). Four of these observers (Observer 2, Observer 4, Observer 8, and Observer 17) met  
39 the performance criterion set for screening (2 Female, 2 Male; age 23-56; mean age 38.25). All observers  
40 had normal or corrected-to-normal vision (20/40 or better in both eyes, assessed using Snellen chart) and  
41 normal color vision (0 Ishihara plates read incorrectly). Observers were dark adapted before performing  
42 the experiments. The choice of four observers to complete the experiment was specified in our pre-  
43 registered protocol.  
44

#### 45 **5.5 Stimulus Design**

46 We measured lightness discrimination thresholds as a function of the amount of variability in the surface  
47 reflectance of the background objects. The reflectances were chosen from a distribution of natural  
48 surfaces. The amount of variability was controlled by multiplying the covariance matrix of the  
49

1 distribution by a scalar (See Methods: Reflectance and Illumination Spectra). We measured thresholds for  
2 six logarithmically spaced values of the covariance scalar.

3 For each value of the covariance scalar, we generated a dataset of 1100 images. The dataset had 100  
4 images each at 11 values of the target object LRF. The LRF of the target object in the standard images  
5 was 0.4 and the lightness in the comparison image varied between 0.35 and 0.45 at steps of 0.01 (11  
6 comparison levels). We generated 100 images at each comparison level, each with a different choice of  
7 the reflectance spectra of the background scene objects. For scale factor 0.00 we generated a set of 11  
8 images, one at each LRF level, as the background remained fixed in this case. All images were generated  
9 without secondary reflections specified in the rendering process. The spectral power distribution of each  
10 light source in the scene was fixed over all images. We chose this to be the standard daylight spectrum  
11 D65 (See Methods: Reflectance and Illumination Spectra). The geometry of the 3D scene was also held  
12 fixed.

13

## 14 **5.6 Experimental Details**

15

16 We used a two-interval forced choice procedure to measure thresholds. We showed two images, one after  
17 the other, on a calibrated computer monitor and asked the observer to report the interval in which the  
18 target object was lighter. We fixed the reflectance of the target object in the standard image and varied the  
19 reflectance of the target object in the other image, which we refer to as the comparison image. The  
20 method of constant stimuli was used. The temporal order in which the standard and comparison images  
21 were presented was randomized on each trial. Audio feedback was provided after every trial, with the  
22 correct response determined by which object had the higher LRF.

23

24 We define a trial as the presentation of the two images (standard and comparison images) and collection  
25 of the observer's response. We define an interval as the presentation of one of the images in the trial.  
26 Thus, a trial has two intervals.

27

28 The experiment was structured as follows. We define an acquisition as the data collected at one  
29 covariance scale factor with 30 trials at each of the 11 comparison levels. We define a permutation as a  
30 set of six acquisitions, where each acquisition corresponds to one of the possible six scale factors. We  
31 collected three permutations for each observer, with a new random order drawn for each permutation.  
32 Thus, after the practice session (see: Recruitment and Exclusion), there were total 18 acquisitions. We  
33 divided these 18 acquisitions over 6 sessions, each session with 3 acquisitions. In each acquisition, we  
34 randomly selected the images on the trials from the pre-generated image databases. The first five trials of  
35 each acquisition were moderate trials (as defined above in Observer Recruitment and Exclusion) to  
36 acclimatize the observer to the experimental task. The responses for these five trials were not saved.

37

38 Each acquisition thus consisted of 330 trials (excluding the 5 moderate acclimatization trials), 30 at each  
39 of the 11 comparison levels. The trial sequence (order of comparison stimuli) in an acquisition was  
40 generated pseudo-randomly at the beginning of the acquisition. For this, at each comparison lightness  
41 level, 30 standard and comparison images were chosen pseudo-randomly with replacement from the  
42 image dataset. The sequence of presentation of these 330 trials were randomized and saved. For each trial,  
43 the order of presentation of the standard and comparison image was also determined pseudo-randomly  
44 and saved. The trials were presented according to the saved sequence.

45

46 The trials in an acquisition were presented in three blocks of 110 trials each. At the end of each block  
47 observer took a rest (of minimum 1 minute). The observer could terminate the experiment anytime during  
48 the acquisition. If an observer terminated an acquisition, the data for that acquisition was not saved. No  
49 observer terminated any acquisition. One observer rescheduled at the beginning of a session due to  
50 tiredness for reasons unrelated to the experiment. The session was rescheduled.

1 At the beginning of the first experimental session (the practice session) for an observer, the experimenter  
2 explained the experimental procedures and obtained consent for the experiments. The experimenter then  
3 tested the observers for normal visual acuity and color vision. The observers were then taken to the dark  
4 room where the observers were described the task and familiarized with the display, chin rest, and  
5 response box. Once familiar, the observers were dark adapted (by sitting in the dark room for  
6 approximately 5 minutes). Once ready, the observers performed the familiarization acquisition. After the  
7 familiarization acquisition, the observers performed the other three acquisitions of the practice session.  
8 The entire practice session took nearly one hour.

9  
10 The observers who met the criteria performed 18 acquisitions over 6 other sessions. The order of these  
11 acquisitions was determined pseudo-randomly at the beginning of the practice session. In each session,  
12 the observer performed only three acquisitions. The observers were dark adapted at the beginning of each  
13 session.

#### 14 **5.7 Stimulus Presentation**

15  
16 The size of each image was 2.6cm x 2.6cm on the monitor, corresponding to  $2^\circ$  by  $2^\circ$  visual angle. The  
17 target object size on the screen in the 2D images was  $\sim 1^\circ$  in diameter. Each image was presented for  
18 250ms (this was a deviation from the preregistration document, which specifies the presentation time as  
19 500ms), with an inter-stimulus interval of 250ms and inter-trial interval of 250ms. Inter-stimulus interval  
20 (ISI) is defined as the interval between the first and the second image presented on each trial. The  
21 response for each trial was collected after both the images had been displayed and removed from the  
22 screen. The observer could take as long as they wished before entering the response. Feedback was  
23 provided via tones presented after the response. The next trial was presented 250ms (ITI) after the  
24 feedback. Thus, the actual inter-trial interval depended on the response time of the observer.  
25

#### 26 **5.8 Image Generation**

27  
28 The images were generated using software we refer to as Virtual World Color Constancy (VWCC)  
29 ([github.com/BrainardLab/VirtualWorldColorConstancy](https://github.com/BrainardLab/VirtualWorldColorConstancy)). VWCC is written using MATLAB. It harnesses  
30 the Mitsuba renderer to render simulated images from scene descriptions, and also takes advantage of our  
31 RenderToolbox package ([rendertoolbox.org](http://rendertoolbox.org); Heasly, Cottaris, Lichtman, Xiao, & Brainard, 2014). To  
32 render an image, we first create a 3D model that specifies the base scene. Objects and light sources can be  
33 inserted in the base scene at user specified locations. The 3D models were based on a base scene provided  
34 as part of RenderToolbox and modified using Blender, an open-source 3-D modeling and animation  
35 package ([blender.org](http://blender.org)). Next, we assigned reflectance spectra and spectral power distribution functions to  
36 the objects and light sources in the scene (see Reflectance and Illumination Spectra Generation for  
37 how these spectra were generated). Once the geometrical and spectral features were specified, we render a  
38 2D multispectral image of the scene using Mitsuba, a physically-realistic open-source rendering  
39 system ([mitsuba-renderer.org](http://mitsuba-renderer.org); Jakob, 2010). The images were rendered at 31 wavelengths equally  
40 spaced between 400nm and 700nm. The images were rendered with the camera field of view of  $17^\circ$  with  
41 an image resolution of 320-pixel by 240-pixels with the target object at the center. A 201-pixel by 201-  
42 pixel area, centered around the spherical target object, was cropped for display on the monitor.  
43

44 To present the multispectral images on the monitor, they were first converted to LMS images using the  
45 Stockman-Sharpe  $2^\circ$  cone fundamentals (T\_cones\_ss2 in the Psychophysics Toolbox). Then the monitor  
46 calibration data and standard methods (Brainard, 1989) were used to convert the LMS images to RGB  
47 images. Finally, a common scaling was applied to all of the images to bring them into the display gamut  
48 of the monitor. The gamma corrected RGB images was presented on the monitor during the experiment.

1    **5.9 Reflectance and Illumination Spectra**

2    The reflectance spectra for the objects were generated using random sampling of datasets of natural world  
3    objects as described in Singh et. al (Singh, Cottaris, Heasly, Brainard, & Burge, 2018). We first  
4    approximated the natural datasets using principal component analysis (PCA). We projected the dataset  
5    along the PCA eigenvectors with the largest 6 eigenvalues. For the reflectance spectrum dataset, these  
6    directions capture more than 90% of the variance. We then approximated the resulting distribution by a  
7    multivariate normal distribution. Reflectance spectra for the objects in the scene were generated using  
8    random sampling from this multivariate normal distribution. The reflectance spectra were constructed as a  
9    linear combination of PCA eigenvectors and the sampled weights. The amount of variation in the surface  
10   reflectance of the background objects was controlled by multiplying the covariance matrix of the  
11   distribution with a scalar. We generated images for six logarithmically spaced values of the covariance  
12   scalar [0, 0.01, 0.03, 0.1, 0.3, 1.0]. We imposed a physical realizability condition on the spectral samples  
13   by ensuring that the reflectance at each spectral frequency was within 0 and 1. Due to this condition, the  
14   variance of the generated spectral samples for some covariance scalars was lower than the variance of the  
15   multi-normal distribution.

16   The power spectrum of the light sources was chosen as standard daylight D65 spectrum. We normalized  
17   the D65 spectrum by its mean power to get the relative spectral shape. This was multiplied by a fixed  
18   scalar with an arbitrarily chosen value of 5 to get the illuminant spectrum. This spectrum was used for all  
19   light sources in the visual scene. We then scaled the rendered image dataset so that the maximum linear  
20   monitor input value (across the entire dataset) needed to render the images on the display was 0.9.

21   **5.10 Psychometric Function**

22   The proportion comparison chosen data was used to obtain the psychometric function for each  
23   acquisition. Each acquisition consisted of 330 trials with 30 trials at each comparison lightness level. At  
24   each lightness level, we recorded the number of times the observers chose the comparison image to be  
25   lighter. The proportion comparison chosen data was fit with a cumulative Gaussian using the Palamedes  
26   toolbox (Prins & Kingdom, 2018). The data was fit to obtain all four parameters of the psychometric  
27   function: threshold, slope, lapse rate and guess rate. While estimating the parameters, the lapse rate was  
28   set equal to the guess rate and was forced to be in the range [0, 0.05]. The model was fit to the data using  
29   maximum likelihood method. The threshold was obtained as the difference between the LRFs at  
30   proportion comparison chosen 0.76 and 0.50 as obtained from the cumulative gaussian fit.

31   **5.11 Signal Detection Theory Model**

32   We developed a model of performance in our task based on signal detection theory (Green, 1996). We  
33   model the visual response to the target object in each image by a univariate internal representation  
34   denoted by the variable  $z$ . This variable depends on the image and is perturbed by noise. We assume that  
35   for any fixed image,  $z$  is a Gaussian distributed random variable whose mean depends on the target object  
36   LRF. For each image, we assume that  $z$  is perturbed on a trial-by-trial basis by independent zero-mean  
37   Gaussian noise, and we assume that the variance this noise is the same for the response to all images. We  
38   refer to the noise that perturbs  $z$  for a fixed image as the internal noise and denote its variance as  $\sigma_i^2$ . For  
39   each trial of the experiment,  $z$  takes on two values,  $z_s$  and  $z_c$ , one for the interval containing the standard  
40   and the other for the interval containing the comparison.

41   If we consider performance for a particular pair of target standard and comparison LRFs, performance  
42   depends both on the difference between the expected values of  $z$  for each pair of LRFs,  $\mu_s$  and  $\mu_c$ , and on  
43   the value of  $\sigma_i^2$ . In our experimental design we have an ensemble of images corresponding to each value

1 of the target sphere LRF. The fact that we draw stochastically from this ensemble on each trial introduces  
 2 additional variability into the value of the decision variable  $z$  that corresponds to a fixed target LRF. We  
 3 call this the external variability, and model it as a Gaussian random variable with zero mean and variance  
 4  $\sigma_e^2$ . We assume that  $\sigma_e^2$  depends on the experimentally chosen covariance scalar, but not on the target  
 5 sphere LRF. Thus, the distributions of  $z_s$  and  $z_c$ , for a particular choice of target standard and comparison  
 6 LRF and covariance scalar, are given by  $P(z_s) = N(\mu_s, \sigma_t)$  and  $P(z_c) = N(\mu_c, \sigma_t)$ . Here  $\mu_s$  is the mean  
 7 value of the internal representation to the standard image and  $\mu_c$  is the mean value of the internal  
 8 representation to the comparison image. The total variance  $\sigma_t$  is given as  $\sigma_t^2 = \sigma_i^2 + \sigma_e^2$ , where  $\sigma_i^2$  and  $\sigma_e^2$   
 9 are the variance of the internal and external noise.

10 In the standard formulation of Signal Detection Theory for a 2AFC task, the observer makes their  
 11 decision based on a comparison of  $z_s$  and  $z_c$ , choosing the interval with the higher value of  $z$ . The  
 12 observer's sensitivity depends on the mean values and the variance and is captured by the quantity  $d'$   
 13 prime given as  $d' = (\mu_c - \mu_s)/\sigma_t$ . This quantity ( $d'$ ) measures the distance between the two distributions  
 14 in standard deviations units. Thus  $d' = 0$  corresponds to an inability to distinguish between the standard  
 15 and the comparison image. Larger values of  $d'$  indicate increasing discriminability.

16 For a fixed value of  $d'$ , the difference in mean values is directly proportional to the standard deviation  $\sigma_t$ :

$$17 \quad (\mu_c - \mu_s) = d' \sigma_t = d' \sqrt{(\sigma_i^2 + \sigma_e^2)} \quad (2)$$

18 We further assume that the difference in mean value of the internal variable ( $\mu_c - \mu_s$ ) is proportional to  
 19 the difference in the LRFs of the target object in the standard and comparison images ( $\Delta_{\text{LRF}}$ ). That is,  
 20  $(\mu_c - \mu_s) = C \Delta_{\text{LRF}}$ , where  $C$  is the proportionality constant. Then we have,

$$22 \quad \Delta_{\text{LRF}} = \frac{d'}{C} \sqrt{(\sigma_i^2 + \sigma_e^2)} \quad (3)$$

23 When we measure threshold in a 2AFC task, we choose a criterion proportional correct and find the  $\Delta_{\text{LRF}}$   
 24 that corresponds to that proportion correct. Our choice of 0.76 corresponds to  $d' = 1$ . In addition we can  
 25 choose  $C = 1$ , in essence setting the units for  $z$  to match those of the experimentally determined target  
 26 LRF.

27 In our experiment, the external variability was induced by changing the surface reflectance of the objects  
 28 in the background. We used a multivariate normal distribution to generate the surface reflectance  
 29 functions of the background objects.<sup>3</sup> To change the amount of external noise, we scaled the variance of  
 30 the multinormal distribution by multiplying its covariance matrix with a scalar. Thus, for our experiments  
 31 we can write:

$$32 \quad \Delta_{\text{LRF}} = \sqrt{\sigma_i^2 + \sigma^2 \times \sigma_{e0}^2} \quad (4)$$

33 where  $\sigma^2$  is the covariance scalar and  $\sigma_{e0}^2$  is the external noise introduced when the ensemble of images  
 34 for each value of target LRF has the reflectance of the background surfaces drawn from our model of  
 35 natural surface reflectances.

---

<sup>3</sup> 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 below.

1  
2 Converting the equation above to the form we use to represent the data, we have

3  $\log(\Delta_{\text{LRF}}^2) = \log(\sigma_i^2 + \sigma^2 \times \sigma_{e0}^2)$  (5)

4 The equation above predicts that the form of threshold  $\log(\Delta_{\text{LRF}}^2)$  as a function of covariance scalar  $\sigma^2$   
5 should increase monotonically. For small values of  $\sigma^2$  ( $\sigma^2 \ll \sigma_i^2/\sigma_{e0}^2$ ), the threshold will approach a  
6 constant giving  $\log(\Delta_{\text{LRF}}^2) \sim \log(\sigma_i^2)$ . For large values of  $\sigma^2$  ( $\sigma^2 \gg \sigma_i^2/\sigma_{e0}^2$ ), the quantity  $\log(\Delta_{\text{LRF}}^2)$  will  
7 approach a straight line with slope 1 in the  $\log(\Delta_{\text{LRF}}^2)$  versus  $\log(\sigma^2)$  plot. Fitting the measurements with  
8 Equation 5 allows us to check whether the model describes the data, as well as, to determine the two  
9 parameters  $\sigma_i^2$  and  $\sigma_{e0}^2$ . In particular, we can establish the relative contribution of the internal  
10 representational variability and external stimulus drive variability in limiting lightness discrimination.  
11 Indeed, the parameter  $\sigma_{e0}^2$  quantifies how much the variation in background surface reflectance intrudes  
12 on the internal representation  $z$  that mediates the lightness discrimination task, in a manner that may be  
13 compared to the intrinsic precision of that representation specified by  $\sigma_i^2$ .

## 14 5.12 Linear Receptive Field Model

15 When the external noise added to the images is characterized by a multivariate Gaussian, a simple linear  
16 receptive field (LINRF) model of the visual system is equivalent to the SDT model developed above. We  
17 first develop this equivalence. The advantage of the receptive field formulation is that it can be  
18 implemented computationally and applied in cases where the external noise is not Gaussian. In our case,  
19 the fact that we truncate surface reflectances to lie between 0 and 1 to satisfy physical realizability means  
20 that the Gaussian characterization is only an approximation, so that adopting the linear receptive field  
21 formulation improves the precision of our modeling. This approach also allows us to incorporate the  
22 Poisson variability of the cone excitations.

23 Denote the stimulus image by the column vector  $I$ , and the receptive field by the column vector  $R$ . The  
24 entries of  $I$  are the radiant power emitted by the monitor at each image location. The entries of  $R$  are the  
25 corresponding sensitivities of the linear receptive field to each entry of  $I$ . The response of the receptive  
26 field is given as  $r_i = R^T I + \eta_i$ , where  $\eta_i$  is a random variable representing a draw of zero mean internal  
27 noise (variance  $\sigma_{ri}^2$ ) in the receptive field response for a fixed image. We assume that  $\sigma_{ri}^2$  is independent  
28 of  $I$ .

29 Denote  $I_{s0}$  and  $I_{c0}$  as the standard and comparison images without external noise. External Gaussian  
30 noise is added to both  $I_{s0}$  and  $I_{c0}$ , with covariance matrix  $\Sigma_e$ . The external noise need not have zero mean.  
31 After incorporation of the external noise, the response of the receptive field to the comparison and  
32 standard images is given by

33  $r_{ic} = R^T(I_{c0} + \eta_e) + \eta_i = R^T I_{c0} + \eta$  (6)

34  $r_{is} = R^T(I_{s0} + \eta_e) + \eta_i = R^T I_{s0} + \eta$ . (7)

35 Here  $\eta_e$  is a random variable representing a draw of external noise,  $\eta_i$  represents the internal noise, and  $\eta$   
36 is a random variable representing the overall effect of the external and internal noise. Since the receptive  
37 field and noise models are linear and Gaussian,  $\eta$  is Gaussian with variance

38  $\sigma_\eta^2 = (\sigma_{ri}^2 + R^T \Sigma_e R)$ . (8)

1 The mean difference between the receptive field response to the comparison and the standard image is  
 2 given by  $(\mu_c - \mu_s) = R^T(I_{c0} - I_{s0}) = C'\Delta_{\text{LRF}}$ . Here  $I_{s0}$  and  $I_{c0}$  are the standard and comparison images  
 3 without external noise added,  $C'$  is a constant, and  $\Delta_{\text{LRF}}$  is as defined in the SDT section above.

4 We associate the linear receptive field response with the internal representation  $z$  of the SDT model  
 5 developed above. That is, we assume that on each trial, the observer chooses as lighter the interval for  
 6 which the response of the receptive field is greater. Following the development of the SDT model, we  
 7 have

$$8 \quad \Delta_{\text{LRF}} = \frac{d'}{C'} \sqrt{\sigma_{ri}^2 + \sigma^2 \times (R^T \Sigma_{e0} R)} \quad (9)$$

9 where we have introduced the covariance scalar  $\sigma^2$  in the term corresponding to the variance of the  
 10 external noise and where  $\Sigma_{e0}$  denotes the covariance matrix of the external noise corresponding to the  
 11 level of variation in natural images. Comparing to relation derived in the SDT model (Equation 4), we see  
 12 that this is the same functional form for the relation between  $\Delta_{\text{LRF}}$  and  $\sigma^2$  as derived there, where we  
 13 associate  $\sigma_i^2 = \frac{\sigma_{ri}^2}{(C')^2}$  and  $\sigma_{e0}^2 = \frac{(R^T \Sigma_{e0} R)}{(C')^2}$ .

14 To fit this model, we use a one-parameter description of a simple center-surround receptive field and use  
 15 simulation to compute model responses for any choice of  $\sigma_i^2$ . This procedure is described in more detail  
 16 below. Once the fitting procedure (described below) establishes  $R$  and  $\sigma_i^2$  that best account for the data,  
 17 we then find  $\sigma_{e0}^2$  directly by passing the images corresponding to  $\sigma^2 = 1$  through the receptive field and  
 18 finding the resulting variance.

## 19 **5.13 SDT Model Fit**

20 The theory of signal detection model was fit to the threshold versus covariance scalar data to obtain the  
 21 parameters  $\sigma_i^2$  and  $\sigma_e^2$ . The parameters were obtained by minimizing the mean squared error between the  
 22 measured and predicted threshold using the MATLAB function *fmincon*.

## 23 **5.14 Linear Receptive Field Model Fit**

24 We fit the linear receptive field (LINRF) model using a simulation approach. We used simulation for two  
 25 reasons. First, it allowed us to incorporate a model of the early visual system into the computations.  
 26 Second, it provides a way to account for truncation in the Gaussian model of natural surface reflectances.  
 27 The truncation occurs because we require that surface reflectance at each wavelength lie between 0 and 1.

28 The model of early visual system was as described by Singh et al. (2018). The model was implemented  
 29 using the software infrastructure provided by ISETBio (ISETBio; [isetbio.org](http://isetbio.org); Cottaris, Jiang, Ding,  
 30 Wandell, & Brainard, 2019). It incorporated typical optical blurring, axial chromatic aberration  
 31 (Marimont & Wandell, 1994), and spatial sampling by the mosaic of long (L), middle (M) and short (S)  
 32 wavelength-sensitive cones (Brainard, 2015). The L:M:S cone ratio in the cone mosaic was chosen to be  
 33 0.6:0.3:0.1 (1523 L cones, 801 M cones, 277 S cones). The CIE physiological standard (CIE, 2007) as  
 34 implemented in ISETBio was used to obtain LMS cone fundamentals. Cone excitations were calculated  
 35 as the number of photopigment isomerizations in a 100ms integration time, and included simulation of the  
 36 Poisson variability of the isomerizations (Rodieck, 1998). The cone isomerizations were demosaiced  
 37 using linear interpolation to estimate LMS isomerization images. Further, the isomerizations of each cone  
 38 class was normalized by the summed (over wavelength) quantal efficiency of the corresponding cone  
 39 class, to make the magnitude of the signals from the three cone classes similar to each other. This

1 normalization occurred after incorporation of Poisson noise and did not affect the signal-to-noise ratio of  
2 the signals from the different cone classes.

3 The dot product of the LMS isomerization images was taken with a simple center-surround linear  
4 receptive field. The receptive field was square in shape to match the image size. Its center was a circle of  
5 radius equal to the size and at the location of the target object in the image. The central region was taken  
6 to have spatially uniform positive sensitivity, while the surround was taken to have spatially uniform  
7 negative sensitivity. Each point in the central region had sensitivity  $v_c = 1$ , and each region of the  
8 surround had sensitivity denoted by  $v_s$ . The RF was the same for each of the three cone classes. The RF  
9 response was taken as the sum of the L, M and S RF component responses. Gaussian internal noise with  
10 zero mean was added to the resulting dot product. The variance of the internal noise ( $\sigma_{ri}$ ) and the value of  
11 the RF surround ( $v_s$ ) were the two parameters of the model.

12 The threshold predictions of the linear receptive field model for any choice of model parameters were  
13 obtained using simulation of a two-interval forced choice paradigm similar to the experiment. For each  
14 trial, we randomly sampled a standard image and a comparison image from our dataset. We obtained the  
15 response of the receptive field (noise-added dot product) to the images and compared them to determine  
16 the simulated choice on that trial. This process was repeated 10,000 times for each of the 11 comparison  
17 LRF levels. The proportion comparison chosen data was used to get the psychometric function and the  
18 threshold of discrimination, similar to the method used for human data. We estimated the threshold at the  
19 six values of covariance scalar at which we performed the human experiments.

20 We calculated the mean squared error (averaged over the six covariance scalar values) between the  
21 thresholds of the human data being fit and the computational model for a large set of values of the two  
22 model parameters: the variance of the decision noise ( $\sigma_{ri}$ ) and the value of the RF surround ( $v_s$ ). The  
23 mean squared error values obtained as a function of these two parameters were fit with a degree two  
24 polynomial of two variables using the MATLAB *fit* function. The resulting polynomial was minimized to  
25 estimate the parameters with lowest mean square error. These parameters were then used to estimate the  
26 internal and external noise standard deviation of the linear receptive field model using the relations:  $\sigma_i^2 =$   
27  $\frac{\sigma_{ri}^2}{(C')^2}$  and  $\sigma_{eo}^2 = \frac{(R^T \Sigma_{eo} R)}{(C')^2}$  as explained above, where the constant  $C'$  was obtained using the relationship  
28  $R^T(I_{c0} - I_{s0}) = C' \Delta_{LRF}$ .

29 The best parameters and the internal and external noise standard deviation were estimated separately for  
30 the mean observer and the individual observers.

### 31 **5.15 Code and Data Availability**

32 Observers' response in the psychophysics task and their thresholds are provided as supplementary  
33 information (SI). The SI also provides the MATLAB scripts to generate Figures 2, 4, 5 and 6 and the  
34 scripts to get thresholds of the linear receptive field model. The retinal images are provided as .mat files  
35 in a zip folder. The SI is available at: <https://github.com/vijaysoophie/EquivalentNoisePaper>

36

1   **APPENDIX**

2   **A1 Measurement of object lightness discrimination thresholds under variation in background**  
3   **object surface reflectance**

4   This supplemental experiment, pre-registered as Experiment 2, provided preliminary data that helped  
5   shape the design of the main experiment presented in the paper (which was Experiment 3 of the pre-  
6   registration documents). It aimed to determine whether variation in the reflectance of background objects  
7   had an effect on human lightness discrimination thresholds. It established that human object lightness  
8   discrimination thresholds increase if the reflectance of background objects vary, as compared to the case  
9   when the discrimination is made against the same background. It also studied the effect of inclusion or not  
10   of secondary reflections in the rendering process as well and assessed the effect of implementing  
11   background variation across trials rather than across intervals.

12   The basic methods were the same as the experiment described in the main paper. The practice session was  
13   conducted with the images in Condition 1 described below. The observers were retained for the  
14   experiment if their average threshold of the last two acquisitions during the practice session was lower  
15   than 0.030. This was a deviation from the pre-registered plan where we set the threshold criterion as  
16   0.025. After collecting data from 8 observers, we realized that the criterion was too strict. Only one  
17   observer had met the criterion. After modifying the threshold criterion, we included two of the initially  
18   discontinued observers in our experiment (Observer 5 and Observer 8). Total of 11 naïve observers  
19   participated in the practice sessions. Four of these observers met the criteria for continuing the  
20   experiment. Two of these observers also participated in the main experiment (Observer 4 and Observer  
21   8).

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

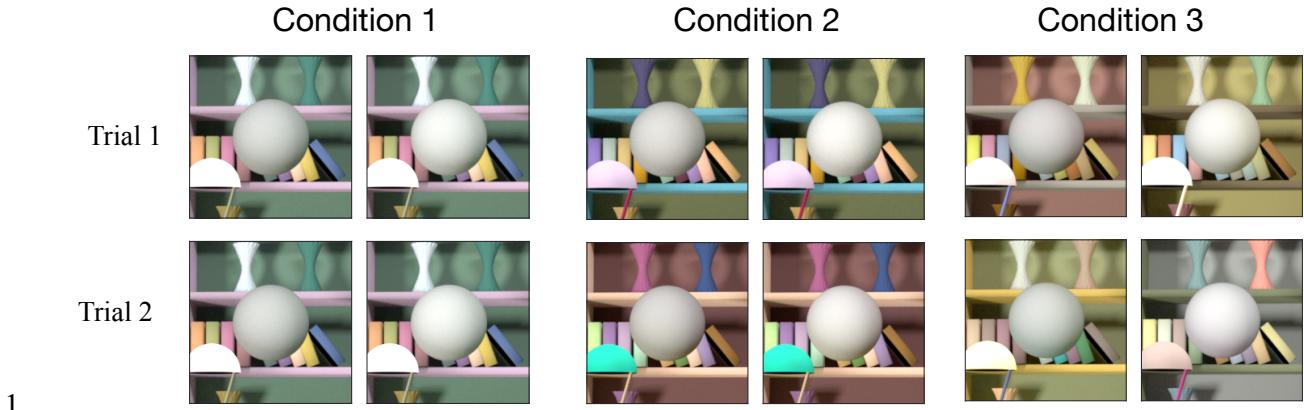
26   *Condition 1. Fixed background:* In this condition, the spectra of objects in the background were kept  
27   fixed for all trials and for all intervals. We generated 11 images, one at each comparison LRF level.

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

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

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

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



2 **Figure S1:** Example stimuli for Conditions 1, 2 and 3 in Experiment 2 to study the effect of background  
 3 surface reflectance on lightness discrimination threshold. In condition 1, the background was fixed in  
 4 every trial and every interval. In condition 2, the background varied from trial to trial, but remained fixed  
 5 in the two intervals of a trial. In condition 3, the background varied in each trial and interval. For  
 6 illustration, in this figure we have chosen the stimulus on the left to be the standard image with target  
 7 object at 0.4 LRF and the on the right to be comparison image with target object at 0.45 LRF. In the  
 8 experiment, the two images were presented sequentially in random order at the center of the screen.  
 9 Conditions 2a and 3a stimuli are similar to condition 2 and 3 respectively, but without secondary  
 10 reflections.  
 11

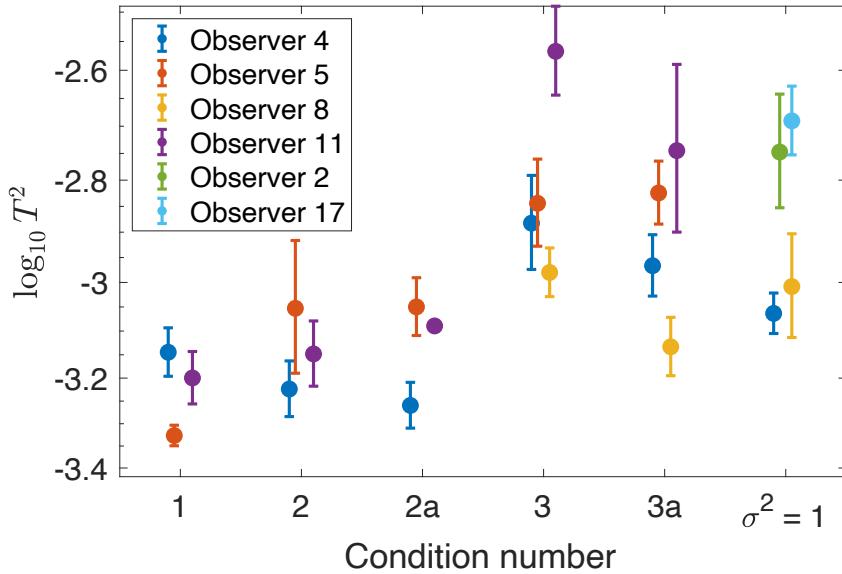
12

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

4    **A2 Results**

5    Figure S2 shows the discrimination thresholds of the four human observers for the five conditions studied  
6    in this experiment. We plot the mean threshold and the standard error of the mean (SEM) taken over the  
7    three separate threshold measurements. For each observer, the thresholds for Condition 3 and 3a were  
8    higher compared to Condition 1, 2 and 2a. The average increases in threshold of the observers for  
9    Conditions 3 and 3a as compared to Condition 1 (baseline) were 79% and 60% respectively. The average  
10   increases in threshold for Conditions 2 and 2a were much smaller, 13% and 17% respectively. The  
11   thresholds for Condition 1, 2 and 2a were nearly within one SEM of each other (averaged over the  
12   observers and three conditions). On the other hand, the thresholds for Conditions 3 and 3a were  
13   respectively (on average) 7.2 and 5.4 SEM larger than the threshold of Condition 1. The thresholds  
14   without secondary reflections (Conditions 2a and 3a) were within one SEM from the conditions with  
15   secondary reflections (Conditions 2 and 3). This preliminary experiment established that lightness  
16   discrimination thresholds are higher for the case when the two objects are being discriminated against  
17   different backgrounds compared on the same trial, as compared to when the backgrounds are the same  
18   within trial. Trial-to-trial variability in background across trials has little, if any, effect. The effect is  
19   similar when the rendering is performed with and without secondary reflections, indicating the effect is  
20   due to the spectral change in the background and not due to the variation in the amount of light being  
21   reflected from the target object surface. In the main experiments, we rendered without secondary  
22   reflections to avoid introducing such variability. Figure S2 also shows the threshold of the observers in  
23   Experiment 3 for the condition with covariance scalar equal to 1. This condition is equivalent to  
24   Condition 3a of Experiment 2. The thresholds of the observers were consistent across the two  
25   measurements.

26



1  
2 **Figure S2:** Lightness discrimination threshold of four human observers in the five conditions in  
3 Experiment 2 (The data points have been jittered to avoid marker overlaps). The thresholds are higher for  
4 the condition where the objects are compared against different backgrounds (Condition 3 and 3a) as  
5 compared to the same background (Condition 1, 2, 2a). Secondary reflections do not have any significant  
6 effect on thresholds (Condition 2a and 3a). Condition 3a of Experiment 2 is equivalent to the condition  
7 with covariance scalar equal to 1 ( $\sigma^2 = 1$ ). The thresholds for this condition are also provided for  
8 comparison. Two observers from Experiment 2 also participated in Experiment 3.  
9

1  
2**Table S1: Observer Thresholds for Experiment 2**

Observer	Mean Threshold +- SEM (averaged over sessions)				
	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

3  
4  
5  
6  
7  
8**Table S2. Lightness discrimination thresholds for Experiment 3:** Mean threshold (averaged over sessions)  
 $\pm$  standard error of measurement of four human observers measured at six logarithmically spaced values of 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

9  
10  
11

1    **REFERENCES**  
2

- 3    Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The  
4    New Cognitive Neurosciences, 2nd edition* (pp. 339-351). Cambridge, MA: MIT Press.  
5    Alvaro, L., Linhares, J. M. M., Moreira, H., Lillo, J., & Nascimento, S. M. C. (2017). Robust  
6    colour constancy in red-green dichromats. *PLoS ONE*, 12(6), e0180310.  
7    American Society for Testing and Materials. (2017). Standard test method for luminous  
8    reflectance factor of acoustical materials by use of integrating-sphere reflectometers.  
9    *Renovations of Center for Historic Preservation*, 98(A), E1477.  
10   Aston, S., Radonjić, A., Brainard, D. H., & Hurlbert, A. C. (2019). Illumination discrimination  
11   for chromatically biased illuminations: implications for colour constancy. *Journal of  
12   Vision*, 19(30:15)  
13   Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility.  
14   *Vision Research*, 27(11), 1915-1924.  
15   Brainard, D. H. (1989). Calibration of a computer controlled color monitor. *Color Research &  
16   Application*, 14(1), 23-34.  
17   Brainard, D. H. (2015). Color and the cone mosaic. *Annual Review of Vision Science*, 1, 519-  
18   546.  
19   Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical  
20   Society of America A*, 14(7), 1393-1411.  
21   Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination  
22   models. *Journal of Vision*, 11(5)  
23   Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. In J. P. Hornak  
24   (Ed.), *Encyclopedia of Imaging Science and Technology* (pp. 172-188). New York: Wiley.  
25   Brainard, D. H., & Radonjić, A. (2014). Color constancy. *The New Visual Neurosciences*, 1,  
26   545-556.  
27   Brascamp, J. W., & Shevell, S. K. (2021). The certainty of ambiguity in visual neural  
28   representations. *Annual Review of Vision Science*, in press  
29   Brindley, G. S. (1960). *Physiology of the Retina and the Visual Pathway*. London: Arnold.  
30   Brown, R. O., & MacLeod, D. I. A. (1997). Color appearance depends on the variance of  
31   surround colors. *Current Biology*, 7, 844-849.  
32   Burge, J. (2020). Image-computable ideal observers for tasks with natural stimuli. *Annual  
33   Review of Neuroscience*, 6, 491-517.  
34   Burge, J., & Geisler, W. S. (2011). Optimal defocus estimation in individual natural images.  
35   *Proceedings of the National Academy of Sciences*, 108(40), 16849-16854.  
36   Burge, J., & Geisler, W. S. (2014). Optimal disparity estimation in natural stereo images. *Journal  
37   of Vision*, 14(2)  
38   Burge, J., & Geisler, W. S. (2015). Optimal speed estimation in natural image movies predicts  
39   human performance. *Nature Communications*, 6, 7900.  
40   Burge, J., & Jaini, P. (2017). Accuracy maximization analysis for sensory-perceptual tasks:  
41   computational improvements, filter robustness, and coding advantages for scaled additive  
42   noise. *PLoS Computational Biology*, 13(2), e1005281.  
43   Chin, B. M., & Burge, J. (2020). Predicting the partition of behavioral variability in speed  
44   perception with naturalistic stimuli. *Journal of Neuroscience*, 40(4), 864-879.

- 1 CIE. (2007). *Fundamental chromaticity diagram with physiological axes – Parts 1 and 2.*  
2      *Technical Report 170-1*. Vienna: Central Bureau of the Commission Internationale de l'  
3      Éclairage.
- 4 Cohen, M. R., & Newsome, W. T. (2009). Estimates of the contribution of single neurons to  
5      perception depend on timescale and noise correlation. *J Neurosci*, 29(20), 6635-6648.
- 6 Cottaris, N. P., Jiang, H., Ding, X., Wandell, B. A., & Brainard, D. H. (2019). A computational-  
7      observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-  
8      mosaic structure, and inference engine. *Journal of Vision*, 19(4), 8.
- 9 Fechner, G. T. (1966). *Elements of Psychophysics*. New York: Holt, Rinehart and Winston.
- 10 Foster, D. H. (2011). Color constancy. *Vision Research*, 51(7), 674-700.
- 11 Gegenfurtner, K., & Kiper, D. C. (1992). Contrast detection in luminance and chromatic noise.  
12      *Journal of the Optical Society of America A*, 9(11), 1880-1888.
- 13 Geisler, W. S., Najemnik, J., & Ing, A. D. (2009). Optimal stimulus encoders for natural tasks.  
14      *Journal of Vision*, 9(13), 17 11-16.
- 15 Gilchrist, A. L. (2006). *Seeing Black and White*. Oxford: Oxford University Press.
- 16 Giulianini, F., & Eskew, R. T., Jr. (1998). Chromatic masking in the (DL/L, DM/M) plane of  
17      cone-contrast space reveals only two detection mechanisms. *Vision Research*, 38, 3913-  
18      3926.
- 19 Green, D. M., & Swets, J. A. (1996). *Signal Detection Theory and Psychophysics* (Vol. 1). New  
20      York: Wiley.
- 21 Heasly, B. S., Cottaris, N. P., Lichtman, D. P., Xiao, B., & Brainard, D. H. (2014).  
22      RenderToolbox3: MATLAB tools that facilitate physically based stimulus rendering for  
23      vision research. *Journal of Vision*, 14(2)
- 24 Helmholtz, H. (1896). *Physiological Optics*. New York: Dover Publications, Inc.
- 25 Henning, G. B., Hertz, B. G., & Hinton, J. L. (1981). Effects of different hypothetical detection  
26      mechanisms on the shape of spatial-frequency filters inferred from masking experiments:  
27      I. Noise masks. *Journal of the Optical Society of America*, 71(5), 574-581.
- 28 Hillis, J. M., & Brainard, D. H. (2005). Do common mechanisms of adaptation mediate color  
29      discrimination and appearance? Uniform backgrounds. *Journal of the Optical Society of*  
30      *America A*, 22(10), 2090-2106.
- 31 Hillis, J. M., & Brainard, D. H. (2007a). Distinct mechanisms mediate visual detection and  
32      identification. *Current Biology*, 17(19), 1714-1719.
- 33 Hillis, J. M., & Brainard, D. H. (2007b). Do common mechanisms of adaptation mediate color  
34      discrimination and appearance? Contrast adaptation. *Journal of the Optical Society of*  
35      *America A*, 24(8), 2122-2133.
- 36 Hurlbert, A. (2019). Challenges to color constancy in a contemporary light. *Current Opinion in*  
37      *Behavioral Sciences*, 30:186, 186-193.
- 38 Ishihara, S. (1977). Tests for colour-blindness. *Tokyo: Kanehara Shuppon Company, Ltd.*
- 39 Jaini, P., & Burge, J. (2017). Linking normative models of natural tasks to descriptive models of  
40      neural response. *Journal of Vision*, 17(12), 16.
- 41 Jakob, W. (2010). Mitsuba Renderer.
- 42 Kelly, K. L., Gibson, K. S., & Nickerson, D. (1943). Tristimulus specification of the Munsell  
43      book of color from spectrophotometric measurements. *Journal of the Optical Society of*  
44      *America*, 33(7), 355-376.

- 1 Kingdom, F. A. (2011). Lightness, brightness and transparency: a quarter century of new ideas,  
2 captivating demonstrations and unrelenting controversy. *Vision Research*, 51(7), 652-  
3 673.
- 4 Knill, D. C., & Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge  
5 University Press.
- 6 Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of*  
7 *the Optical Society of America A*, 4(2), 391-404.
- 8 Losada, M. A., & Mullen, K. T. (1995). Color and luminance spatial tuning estimated by noise  
9 masking in the absence of off-frequency looking. *Journal of the Optical Society of*  
10 *America A*, 12(2), 250-260.
- 11 Lotto, R. B., & Purves, D. (1999). The effects of color on brightness. *Nature Neuroscience*,  
12 2(11), 1010-1014.
- 13 Marimont, D. H., & Wandell, B. A. (1994). Matching color images: the effects of axial  
14 chromatic aberration. *Journal of the Optical Society of America A*, 11(12), 3113-3122.
- 15 Monaci, G., Menegaz, G., Süssstrunk, S., & Knoblauch, K. (2004). Chromatic contrast detection  
16 in spatial chromatic noise. *Visual Neuroscience*, 21, 291-294.
- 17 Murray, R. F. (2021). Lightness perception in complex scenes. *Annual Review of Vision Science*,  
18 *in press*
- 19 Nachmias, J. (1999). How is a grating detected on a narrowband noise masker? *Vision Research*,  
20 39(6), 1133-1142.
- 21 Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: discrimination may be better than  
22 detection. *Vision Research*, 14(10), 1039-1042.
- 23 Nienborg, H., Cohen, M. R., & Cumming, B. G. (2012). Decision-related activity in sensory  
24 neurons: correlations among neurons and with behavior. *Annu Rev Neurosci*, 35, 463-  
25 483.
- 26 Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of  
27 perception. *Annual Review of Neuroscience*, 21(1), 227-277.
- 28 Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. (2014). Chromatic  
29 illumination discrimination ability reveals that human colour constancy is optimised for  
30 blue daylight illuminations. *PLoS ONE* 9(2:e87989), e87989.
- 31 Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding*  
32 *and Efficiency* (pp. 3-24).
- 33 Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A*,  
34 16(3), 647-653.
- 35 Prins, N., & Kingdom, F. A. A. (2018). Applying the model-comparison approach to test specific  
36 research hypotheses in psychophysical research using the Palamedes toolbox. *Frontiers*  
37 *in Psychology*, 9, 1250.
- 38 Radonjić, A., & Brainard, D. H. (2016). The nature of instructional effects in color constancy. *J*  
39 *Exp Psychol Hum Percept Perform*, 42(6), 847-865.
- 40 Radonjić, A., Ding, X., Krieger, A., Aston, S., Hurlbert, A. C., & Brainard, D. H. (2018).  
41 Illumination discrimination in the absence of a fixed surface-reflectance layout. *Journal*  
42 *of Vision*, 18(5:11)
- 43 Radonjić, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N. P., Brainard, D. H., &  
44 Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal*  
45 *of Vision*, 16(11:2), 1-18.
- 46 Rodieck, R. W. (1998). *The First Steps in Seeing*. Sunderland, Mass.: Sinauer.

- 1 Rovamo, J., Franssila, R., & Nasanen, R. (1992). Contrast sensitivity as a function of spatial  
2 frequency, viewing distance and eccentricity with and without spatial noise. *Vision  
3 Research*, 32(4), 631-637.
- 4 Rovamo, J., Raninen, A., & Donner, K. (1999). The effects of temporal noise and retinal  
5 luminance on foveal flicker sensitivity. *Vision Research*, 39, 533-539.
- 6 Ruff, D. A., Ni, A. M., & Cohen, M. R. (2018). Cognition as a window into neuronal population  
7 space. *Annual Review of Neuroscience*, 41, 77-97.
- 8 Sankeralli, M. J., & Mullen, K. T. (1997). Postreceptoral chromatic detection mechanisms  
9 revealed by noise masking in three-dimensional cone contrast space. *Journal of the  
10 Optical Society of America A*, 14(10), 2633-2646.
- 11 Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational  
12 analysis of the relationship between neuronal and behavioral responses to visual motion.  
13 *Journal of Neuroscience*, 16, 1486-1510.
- 14 Singh, V., Cottaris, N. P., Heasly, B. S., Brainard, D. H., & Burge, J. (2018). Computational  
15 luminance constancy from naturalistic images. *Journal of Vision*, 18(13), 19.
- 16 Smithson, H. E. (2005). Sensory, computational, and cognitive components of human color  
17 constancy. *Philosophical Transactions of the Royal Society of London. Series B*,  
18 360(1458), 1329-1346.
- 19 Teller, D. Y. (1984). Linking propositions. *Vision Research*, 24(10), 1233-1246.
- 20 Vrhel, M. J., Gershon, R., & Iwan, L. S. (1994). Measurement and analysis of object reflectance  
21 spectra. *Color Research & Application*, 19(1), 4-9.
- 22 Weiss, D., Witzel, C., & Gegenfurtner, K. (2017). Determinants of colour constancy and the blue  
23 bias. *i-Perception*, 8(6), 204166951773963.
- 24 Westheimer, G. (1986). The eye as an optical instrument. In K. R. Boff, L. Kaufman & J. P.  
25 Thomas (Eds.), *Handbook of Perception and Human Performance: Sensory Processes  
26 and Perception* (pp. 4.1-4.20). New York: John Wiley & Sons.
- 27 Witzel, C., & Gegenfurtner, K. R. (2018). Color perception: objects, constancy, and categories.  
28 *Annual Review of Vision Science*, 4, 475-499.
- 29 Zhang, X., & Brainard, D. H. (2004). *Bayesian color correction method for non-colorimetric  
30 digital image sensors*. Paper presented at Color and Imaging Conference, 308-314.
- 31