

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>2,3,4,5</sup>

3 1 Department of Physics, North Carolina Agricultural and Technical State University, Greensboro, NC,  
4 USA.

5 2 Computational Neuroscience Initiative, University of Pennsylvania, Philadelphia, PA, USA.

6 3 Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA.

7 4 Neuroscience Graduate Group, University of Pennsylvania, Philadelphia, PA, USA.

8 5 Bioengineering Graduate Group, University of Pennsylvania, Philadelphia, PA, USA.

9

10

11 **ABSTRACT:** A goal of visual perception is to provide stable representations of task-relevant scene  
12 properties (e.g. object reflectance) despite variation in task-irrelevant scene properties (e.g. illumination,  
13 reflectance of other nearby objects). To study such stability in the context of the perceptual representation  
14 of lightness, we introduce a threshold-based psychophysical paradigm. We measure how thresholds for  
15 discriminating the achromatic reflectance of a target object (task-relevant property) in rendered  
16 naturalistic scenes are impacted by variation in the reflectance functions of background objects (task-  
17 irrelevant property). We refer to these thresholds as lightness discrimination thresholds. We measure them  
18 using a two-alternative forced-choice paradigm in which the reflectance of the background objects is  
19 randomized across the two intervals of each trial. We parametrically scale the amount of background  
20 reflectance variation by manipulating a statistical model of naturally-occurring surface reflectance. This  
21 approach has roots in the equivalent noise paradigm, in which effects of internal and external sources of  
22 noise on contrast coding are inferred from threshold measurements. Here, we investigate the interaction of  
23 distinct distal-stimulus variables for lightness coding. For low background object reflectance variation,  
24 lightness discrimination thresholds were nearly constant, indicating that observers' internal noise  
25 determines threshold in this regime. As background object reflectance variation increases, its effects start  
26 to dominate performance. A model based on signal detection theory, that utilizes the responses of a linear  
27 receptive field, captures human behavior in this task and allows us to express the effects of task-irrelevant  
28 variation relative to the intrinsic precision of the task-relevant perceptual representation. The results  
29 motivate the conclusion that naturally occurring background object reflectance variation corrupts the  
30 perceptual representation of lightness. For our stimulus conditions, however, the threshold increases  
31 resulting from background object reflectance variation are modest - within a factor of two of the limit set  
32 by internal noise. Our approach introduces a threshold-based method for characterizing the effect of task-  
33 irrelevant scene variation (i.e. external variability) on the perceptual representation of a task-relevant  
34 scene property.

35 **KEYWORDS:** Lightness, Noise Masking, Equivalent Noise, Human Psychophysics, Color Vision

36

37 **INTRODUCTION**

38  
39 To support effective action, vision provides stable perceptual representations of the distal properties of  
40 objects. The computations that give rise to these representations start with the information in the proximal  
41 stimuli that are encoded by the retinas. These proximal stimuli depend on the intrinsic properties of the  
42 objects in the scene, on object-extrinsic properties of the scene (e.g. illumination), and on the observer's  
43 particular viewpoint. A challenge for the visual system is to recover stable perceptual correlates of object-  
44 intrinsic properties across variation in other scene variables. Understanding the degree to which the visual  
45 system rises to this challenge, and how it does so, is an important goal of vision science (Helmholtz,  
46 Knill & Richards, 1996; Brascamp & Shevell, 2021; Geisler, 2008; Burge, 2020; Wandell &  
47 Brainard, in press).

48  
49 Here we consider the perceptual task of representing the reflectance of an object embedded in a scene,  
50 based on the light reflected to the eye from the object and the rest of the scene. The perceptual correlate of  
51 object surface reflectance is its perceived color or, in the special case of achromatic objects, its lightness.  
52 Computing a stable color or lightness representation poses a challenge to the visual system because the  
53 retinal image of the object varies with the object's reflectance, the spectral irradiance of the illumination,  
54 the position and pose of the object in the scene, and the properties of other objects in the scene. The  
55 degree to which the visual system succeeds at stabilizing its color and lightness representations of objects,  
56 in the face of variation extrinsic to their reflectance, determines the degree to which the visual system  
57 achieves color and lightness constancy.

58  
59 Under many circumstances, the visual system achieves a high degree of color and lightness constancy  
60 (Foster, 2011). Several theoretical frameworks have been developed to account for this ability. The  
61 frameworks attempt to explain how different cues are processed to form stable perceptual representations  
62 of object reflectance (Adelson, 2000; Smithson, 2005; Gilchrist, 2006; Kingdom, 2011; Foster, 2011;  
63 Brainard & Maloney, 2011; Brainard & Radonjić, 2014; Witzel & Gegenfurtner, 2018; Hurlbert, 2019;  
64 Murray, 2021). The underlying computations have been explained in terms of mechanistic gain control  
65 (e.g., von Kries, 1905; Whittle & Challands, 1969; Webster & Mollon, 1995; Land & McCann, 1971;  
66 Horn, 1974), cue combination (e.g., Maloney & Yang, 2001; Yang & Maloney, 2001), Bayesian inference  
67 (e.g., Brainard & Freeman, 1997; Brainard et al., 2006; Barron & Malik, 2012a; Allred & Brainard, 2013;  
68 Murray, 2020; see also Boyaci, Maloney, & Hersh, 2003; Bloj et al., 2004; Brainard & Maloney, 2011),  
69 learned computations (e.g., Singh, Cottaris, Heasly, Brainard, & Burge, 2018; Flachot & Gegenfurtner,  
70 2018; Flachot & Gegenfurtner, 2021; Afifi, Barron, LeGendre, Tsai, & Bleibel, 2021), and application of  
71 principles of perceptual organization (Gilchrist, 2006; Adelson, 1993).

72  
73 Color constancy and lightness constancy have been elucidated primarily with an experimental approach in  
74 which observers report on suprathreshold aspects of the color or lightness of a target object, across  
75 changes extrinsic to the target object's reflectance.<sup>1</sup> In these experiments, the target object's reflectance is  
76 the task-relevant scene variable, while other aspects of the scene are task irrelevant. Observers' reports  
77 are solicited using a variety of methods, including matching (e.g., Burnham, Evans, & Newhall, 1952;  
78 Gilchrist, 1977; Arend & Reeves, 1986; Brainard, Brunt, & Speigle, 1997), naming (e.g., Helson &  
79 Jeffers, 1940; Olkkonen, Witzel, Hansen, & Gegenfurtner, 2010), scaling (e.g., Schultz, Doerschner, &  
80 Maloney, 2006), and nulling (e.g., Helson & Michels, 1948; Jameson & Hurvich, 1955; Chichilnisky &  
81 Wandell, 1997; Brainard, 1998).

---

<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 some aspect 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.

82  
83 In the study of perception, discrimination experiments complement experiments that rely on  
84 suprathreshold reports. In a typical discrimination experiment, observers choose which of two stimuli has  
85 a larger physical value along some stimulus dimension. The stimulus difference is titrated to determine  
86 the smallest change that supports criterion discrimination performance. This smallest change is defined as  
87 threshold. For example, observers might be tasked with reporting which of two objects has a larger  
88 lightness value, in an effort to determine the human ability to discriminate different object surface  
89 reflectances. Mature theory links discrimination thresholds to the precision of the underlying perceptual  
90 representation (Green, 1966). Theory also exists for linking thresholds to properties of neural responses  
91 (Brindley, 1960; Green, 1966; Teller, 1984; Parker & Newsome, 1998).

92  
93 Theory is less well developed for how to use discrimination experiments to address questions about  
94 perceptual constancy. In the case of color constancy, one approach is to measure observers' ability to  
95 discriminate changes in scene illumination (Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert, 2014;  
96 Radonjić et al., 2016; Radonjić et al., 2018; Aston, Radonjić, Brainard, & Hurlbert, 2019; Alvaro,  
97 Linhares, Moreira, Lillo, & Nascimento, 2017), rather than to measure the ability to detect a change in  
98 object surface reflectance per se (for work that measures reflectance discrimination thresholds see  
99 Morimoto & Smithson, 2018). The idea is that if illumination changes are subthreshold, then the  
100 perceptual representations of both surface reflectance and illumination are stable across those illumination  
101 changes. However, it is unclear how the results of these experiments connect to and inform us about the  
102 stability of perceptual judgments across the larger illumination changes that occur in natural viewing (but  
103 see Weiss, Witzel, & Gegenfurtner, 2017). Another approach is to link discrimination thresholds to  
104 suprathreshold reports of perceived stimulus properties, an approach which has its origins in Fechner's  
105 pioneering interpretation of Weber's Law (Fechner, 1860). The idea is that both threshold and  
106 suprathreshold percepts are mediated by a common stimulus-response function whose properties depend  
107 on, and change with, viewing context. Although positing a common stimulus-response function holds  
108 promise (Nachmias & Sansbury, 1974; Hillis & Brainard, 2005; Hillis & Brainard, 2007b), there are cases  
109 in which the discrimination thresholds do not predict suprathreshold measures of lightness constancy  
110 made using well-matched stimuli (Hillis & Brainard, 2007a).

111  
112 Here, we introduce a new approach to using discrimination experiments to study perceptual constancy.  
113 The approach is based on measuring how discrimination thresholds for a task-relevant scene property are  
114 affected by variation in a task-irrelevant scene property. The approach is conceptually similar to studying  
115 how contrast thresholds are affected by addition of random, unpredictable stimulus variation, usually  
116 introduced in the form of spatially white or pink contrast noise (Legge, Kersten, & Burgess, 1987; Pelli,  
117 1990; Pelli & Farell, 1999). It is conceptually distinct in that the random, unpredictable variation is  
118 introduced in the distal scene properties (for related recent work, see Zhu, Yuille, & Kersten, 2021). We  
119 apply this approach to the study of lightness constancy in naturalistic scenes. First, we measure human  
120 ability to discriminate the achromatic surface reflectance of a target object in the absence of any target  
121 object-extrinsic variation. Next, we measure how these lightness discrimination thresholds change with  
122 the introduction of target object-extrinsic variation. Specifically, we introduce random, unpredictable  
123 variation to the background objects in the scene by varying their reflectance spectra—loosely, their colors  
124 (Lotto & Purves, 1999; Brown & MacLeod, 1997). The lightness discrimination threshold at each level of  
125 background object reflectance variation measures how difficult the lightness discrimination task is for that  
126 level of variation. The change in difficulty from baseline (i.e., no background object reflectance variation)  
127 quantifies the degree to which the background variation intrudes on the perceptual representation of target  
128 lightness.

129  
130 As the variation in background object reflectances is increased, we find that discrimination thresholds are  
131 initially constant and then increase. To interpret these findings, we develop a model based on signal  
132 detection theory; the model is similar to those used to understand the effect of contrast noise on contrast

133 thresholds (Legge, Kersten, & Burgess, 1987; Pelli, 1990). The model relates thresholds for the task-  
134 relevant variable (here, target object reflectance) to the amount of variation in the task-irrelevant variable  
135 (here, background object reflectance). The model allows us to express the effect of task-irrelevant  
136 variation in terms of equivalent noise. Equivalent noise is the amount of external task-irrelevant variation  
137 whose effect on the perceptual representation is the same as that of internal noise. We find that the  
138 intrusion of naturally occurring variation in background object reflectances on the perceptual  
139 representation of lightness is within a factor of two of the equivalent noise.

140  
141 The paper is organized as follows: Section 2 (Methods) provides the experimental methods. Section 3  
142 (Model) introduces the model used to interpret the data, and discusses the concept of equivalent noise in  
143 more detail. Section 4 (Results) reports the experimental results in the context of the model. Section 5  
144 (Discussion) provides a summary. The Appendix describes a control experiment and provides  
145 supplemental figures and tables. Additional supplementary information is available online as indicated in  
146 Methods: Code and Data Availability.

## 147 **2 EXPERIMENTAL METHODS**

### 148 **Overview**

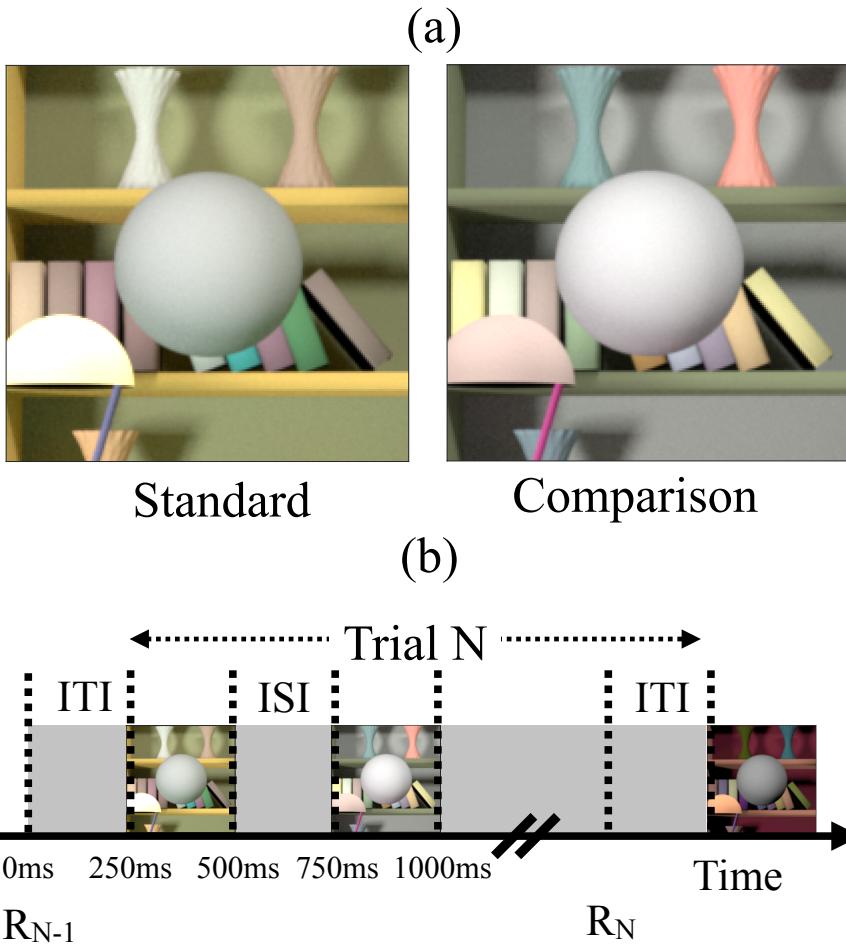
149 We studied the effect of variability in object-extrinsic properties on the human ability to discriminate an  
150 object-intrinsic property. Specifically, we measured how variation in the reflectance spectra of  
151 background objects affects lightness discrimination thresholds, that is thresholds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

164  
165 We recorded the proportion of times observers chose the comparison image as having the lighter target  
166 object at 11 values of the target object LRF. Figure 2 shows a psychometric function from a typical  
167 human observer. The proportion-comparison-chosen data were fit with a cumulative normal using  
168 maximum likelihood methods (see Methods: Psychometric Function). Threshold was defined as the  
169 difference between the LRF of the target object at proportion comparison chosen 0.76 and 0.50 (i.e., d-  
170 prime = 1.0 in a two-interval task), as determined from the cumulative normal fit.

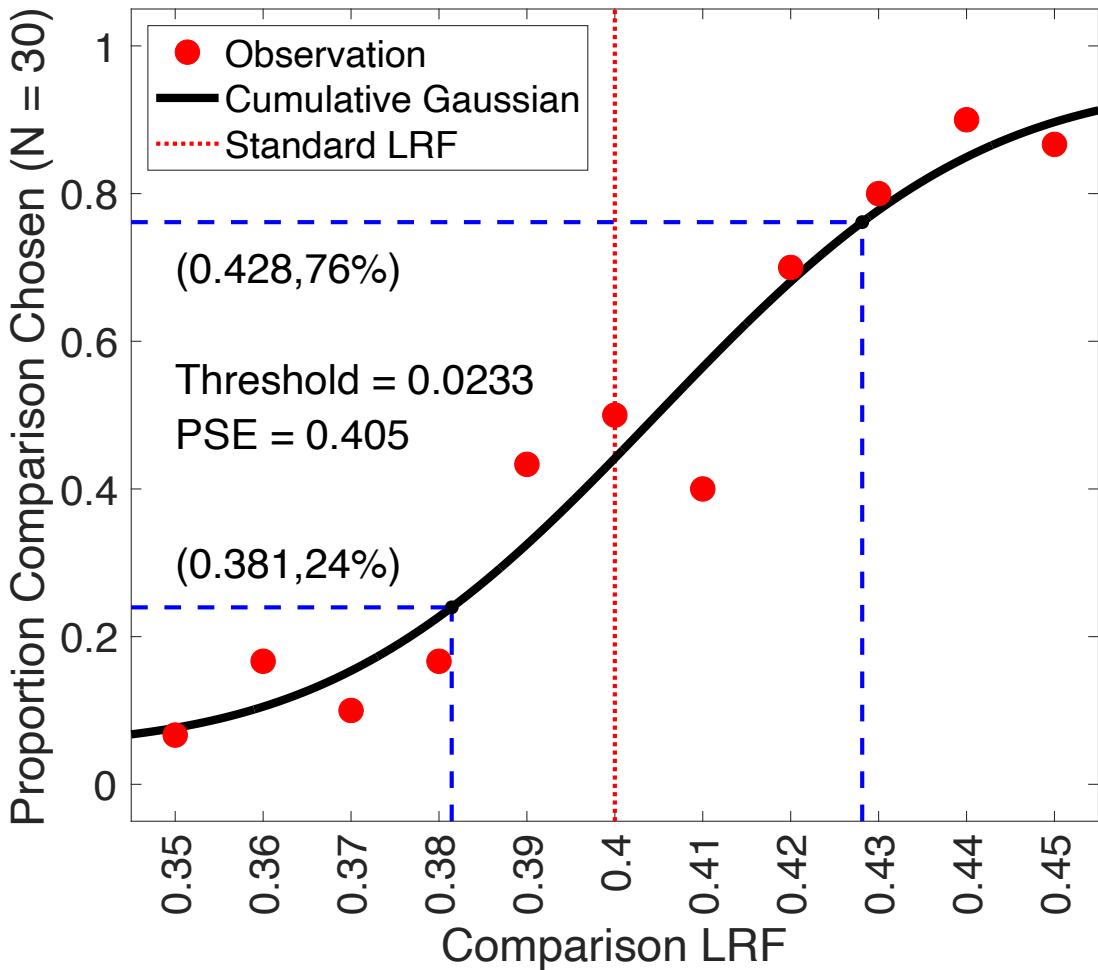
---

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

<sup>3</sup> We use LRF rather than the more generic term albedo as our single number summary of the underlying spectral surface reflectance function, as the LRF is explicit about how variation in reflectance over wavelength should be taken into account.



**Figure 1: Psychophysical task.** (a) On every trial of the experiment, human observers viewed two images in sequence, a standard image and a comparison image and indicated the one in which the spherical target object in the center of the image was lighter. Example standard and comparison images are shown. The images were computer graphics simulations. The simulated reflectance functions of the target were spectrally flat, and the spheres appeared gray. The overall reflectance of the target was held fixed in the standard images and differed between standard and comparison. Performance (proportion correct) was measured as a function of this difference to determine discrimination threshold. The reflectance spectra of objects in the background could be held fixed or vary between standard and comparison on each trial (as illustrated here). The order of presentation of the standard and comparison images was randomized from trial to trial. Discrimination thresholds were measured as function of the amount of variation in background object reflectances. (b) Trial sequence.  $R_{N-1}$  indicates the time of the observer's response for the  $(N-1)^{\text{th}}$  trial. The  $N^{\text{th}}$  trial begins 250ms after that response (Inter Trial Interval, ITI). The  $N^{\text{th}}$  trial consists of two 250ms stimulus presentation intervals with a 250ms inter-stimulus interval (ISI). The observer responds by pressing a button on a gamepad after the second stimulus has been shown. The observer can take as long as he or she wishes before making the response, with an example response time denoted by  $R_N$  in the figure. The next trial begins 250ms after the response.



**Figure 2: Psychometric function.** We recorded the proportion of times the observer chose the target in the comparison image to be lighter, as a function of the comparison LRF. The LRF of the target object in the standard image was fixed at 0.4. The LRF of the target object in the comparison image were chosen from 11 linearly spaced values in the range [0.35, 0.45]. In each block, thirty trials were presented at each comparison LRF value. We fit a cumulative normal distribution to the proportion comparison chosen data using maximum likelihood methods. The guess and lapse rates were constrained to be equal and were restricted to be in the range [0, 0.05]. The threshold was measured as the difference between the LRF at proportion comparison chosen equal to 0.76 and 0.5, as predicted by the cumulative normal fit. This figure shows the data for Observer 2 for scale factor 0.00, for the block run in the first experimental session for that observer. The point of subjective equality (PSE, the LRF corresponding to proportion chosen 0.5) was close to 0.4 as expected and the threshold was 0.0233. The lapse rate for this fit was 0.05.

171  
172 We measured lightness discrimination thresholds as a function of the amount of variability in the surface  
173 reflectances of the background objects in the rendered scenes. The reflectances of the background objects  
174 were chosen from a distribution of natural reflectances. The amount of variability was controlled  
175 parametrically by multiplying the covariance matrix of the distribution by a scalar (see Methods:  
176 Reflectance and Illumination Spectra). We measured thresholds for six logarithmically spaced values of  
177 this covariance scalar. By varying the scalar from 0 (no variation) to 1 (natural-scene-typical variation),  
178 we examined how background variation affects performance in the task. Figure 3 shows examples of  
179 images used in our psychophysical task for different choices of the covariance scalar.

180 The subsections below provide additional methodological detail.

## 181 **Preregistration**

182 The experimental design and the method for extracting threshold from the data were preregistered before  
183 the start of the experiment. The preregistration documents are publicly available at: <https://osf.io/7tgy8/>.<sup>4</sup>

184 We preregistered three experiments. The first experiment (preregistered as Experiment 1) was abandoned  
185 because the task was too difficult. The findings of the second experiment (preregistered as Experiment 2  
186 and referred to here as the control experiment) provide control data and are reported in the Appendix. In  
187 the body of the paper, we report preregistered Experiment 3 (referred to here as the main experiment).  
188 The details of the experimental methods below refer to preregistered Experiment 3; the methods for  
189 preregistered Experiment 2 were essentially the same with key differences (primarily the conditions  
190 studied) described in the Appendix.

191 A deviation from the preregistered plan for preregistered Experiment 2 was the change in the criteria to  
192 select observers for the experiment. The preregistered criterion for selecting an observer for this  
193 experiment was that an observer would be excluded if their mean threshold for the last two blocks in the  
194 practice session exceeded 0.025. After collecting data from 8 naive observers, we concluded that this  
195 criterion was too strict as only one observer met the criterion. Hence, we increased exclusion threshold  
196 from 0.025 to 0.030. The preregistered plans also indicated that each image would be presented for  
197 500ms, but in the event we shortened this to 250ms.

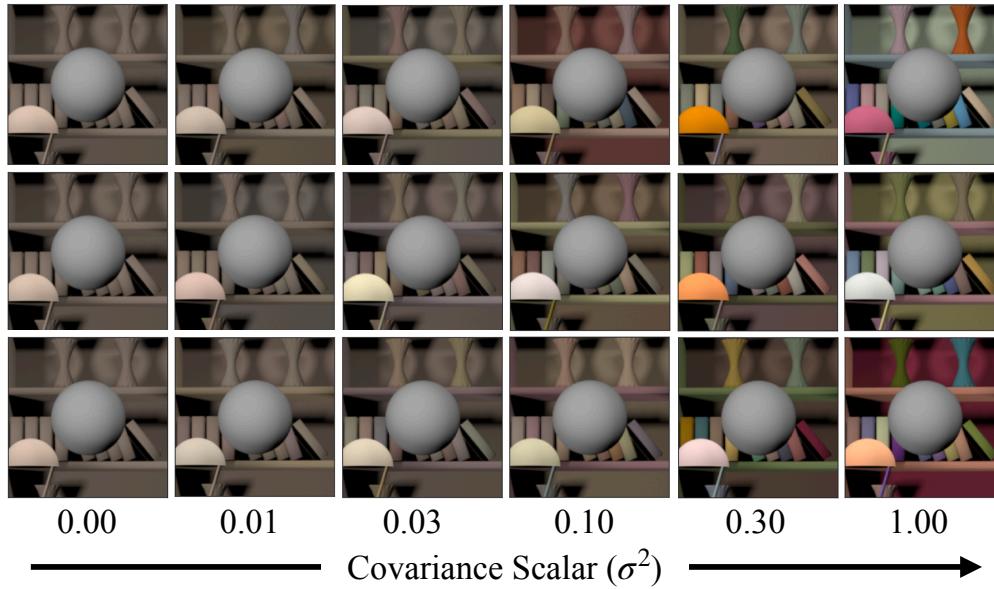
198 We followed the procedure described in the preregistration document to extract threshold from the data.  
199 The document also indicated that the primary data feature of interest was the dependence of threshold on  
200 the covariance scalar and predicted that thresholds would increase with increasing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

## 202 **Reflectance and Illumination Spectra**

203 The reflectance spectra for the background objects in the scene were generated using a model of naturally  
204 occurring surface reflectance spectra, as described in (Singh, Cottaris, Heasly, Brainard, & Burge, 2018).  
205 Briefly, we started with two datasets of surface reflectance functions (Kelly, Gibson, & Nickerson, 1943;  
206 Vrhel, Gershon, & Iwan, 1994) containing 632 surface reflectance measurements in total. The Kelly et al.  
207 dataset has 462 spectral measurements of Munsell papers, with each spectrum available to us  
208 (psychtoolbox.org) on wavelength support 400nm to 700nm at 5nm spacing. The Vrhel dataset has 170  
209 spectral measurements, each spectrum measured in the wavelength range 390nm to 730nm at 2nm

---

<sup>4</sup> The preregistration documents relevant to this paper are those for Experiments 1, 2 and 3. The site also contains preregistrations for subsequent work not reported in this paper.



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

spacing. We converted to a common wavelength support of 10 nm spacing between 400nm and 700nm and combined the two datasets. We then used principal component analysis (PCA) to characterize the combined dataset. For this analysis, we mean centered the dataset and used the singular value decomposition (SVD) to obtain the eigenvectors of the mean-centered dataset. The reflectances in the mean-centered dataset were projected onto the eigenvectors to obtain their projection weights. The eigenvectors associated with the six largest eigenvalues captured more than 99.5% of the variance, so the rest of the analysis focuses on the projection weights on these eigenvectors. We approximated the empirical distribution of projection weights by a multivariate normal distribution. Reflectance spectra for the objects in the scene were generated by randomly sampling from this multivariate normal distribution and using the eigenvectors to construct samples of mean-centered surface reflectances. To these, we added back the mean of the surface reflectance dataset. We imposed a physical realizability constraint on the randomly generated spectral samples by ensuring that the reflectance at each wavelength was between 0 and 1. If the reflectance of a generated sample did not fall in this range at any wavelength, it was discarded.

The amount of variation in the surface reflectance of the background objects was controlled by multiplying the covariance matrix of the multivariate normal distribution (see above) by a covariance scalar. A covariance scalar of 0 corresponds to no background object reflectance variation. A covariance scalar of 1 corresponds to the full reflectance variation of the model of natural reflectance (Figure 3). We generated images for six logarithmically spaced values of covariance scalar: [0, 0.01, 0.03, 0.1, 0.3, 1.0]. Due to the physical realizability constraint, the actual variances of the projection weights for the generated spectral samples for some covariance scalars were lower than the corresponding variances of the underlying multivariate normal, and their distribution was not precisely multivariate normal.

The power spectrum of the light sources was chosen as that of standard daylight D65. We normalized the D65 spectrum by its mean power to obtain its relative spectral shape. This was multiplied by a fixed scalar with an arbitrarily chosen value of 5 to get the illuminant spectrum. This spectrum was used for all light sources in the visual scene and was not varied across the experiments reported here.

## Image Generation

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

Once the geometrical and spectral features were specified, we rendered a 2D multispectral image of the scene using Mitsuba, a physically-realistic open-source rendering system ([mitsuba-renderer.org](http://mitsuba-renderer.org); Jakob, 2010). The images were rendered at 31 wavelengths equally spaced between 400nm and 700nm. The images were rendered with the camera field of view of 17° with an image resolution of 320-pixel by 240-

258 pixels with the target object at the center. A 201-pixel by 201-pixel area, centered around the spherical  
259 target object, was cropped for display on the monitor.  
260

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

## 268 **Stimulus Design**

269 As noted above, we measured lightness discrimination thresholds for six values of the covariance scalar.  
270 For each value of the covariance scalar, we generated a dataset of 1100 images. The dataset had 100  
271 images each at 11 values of the target object LRF. The LRF of the target object in the standard images  
272 was 0.4 and the LRF in the comparison image varied between 0.35 and 0.45 at steps of 0.01 (11  
273 comparison levels). We generated 100 images at each comparison level, each with a different choice of  
274 the reflectance spectra of the background objects. The fact that we had 100 images for each target LRF  
275 allowed us to randomize the background object reflectances across the two intervals of each forced choice  
276 trial without excessive replication. For covariance scalar 0.00 we generated a set of 11 images, one at  
277 each LRF level, as the background remained fixed in this case. All images were generated without  
278 secondary reflections specified in the rendering process. The geometry of the 3D scene was also held  
279 fixed across all images.  
280

281 When displayed on the experimental monitor, the average luminance of the standard image for covariance  
282 scalar 0.00 was 47.3 cd/m<sup>2</sup>. The average luminances of the target object for the 11 LRF levels were [67.0,  
283 68.0, 68.9, 69.8, 70.7, 71.6, 72.5, 73.4, 74.2, 75.1, 75.9] cd/m<sup>2</sup>.  
284

## 285 **Experimental Details**

286 We define a trial as the presentation of two images (standard and comparison images) and collection of  
287 the observer's response. We define an interval as the presentation of one of the images in the trial.  
288

289 The experiment was structured as follows. We define a block of trials as the data collected at one  
290 covariance scalar with 30 trials at each of the 11 comparison levels. We define a permutation as a set of  
291 six blocks, where each block corresponds to one of the possible six covariance scalars. We collected three  
292 permutations for each observer, with a new random order drawn for each permutation. Thus, after the  
293 practice session (see Methods: Observer Recruitment and Exclusion), there were total 18 blocks. We  
294 divided these 18 blocks over 6 sessions, each session with 3 blocks. In each block, we randomly selected  
295 the images for the trials from the pre-generated image database. The first five trials of each block were  
296 moderate trials (as defined in Methods: Observer Recruitment and Exclusion) to acclimatize the observer  
297 to the experimental task. The responses for these five trials were not saved.  
298

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

305  
306 The trials in a block were presented in three sub-blocks of 110 trials each. At the end of each sub-block  
307 the observer took a break of minimum duration 1 minute. The observer could terminate the experiment  
308 anytime during the block. If an observer terminated a block, the data for that block was not saved. No  
309 observer terminated any block. One observer indicated a desire to postpone at the beginning of a session,  
310 due to fatigue for reasons unrelated to the experiment. The session was rescheduled.  
311

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

321 Observers who met the inclusion criteria (see Methods: Observer Recruitment and Exclusion) then  
322 performed 18 blocks over 6 additional sessions, each on a separate day. The order of blocks for each  
323 observer was determined pseudo-randomly at the beginning of the practice session. As noted above,  
324 observers performed three blocks per session. Observers were dark adapted for 5 minutes at the beginning  
325 of each session. The data for all observers in the main experiment (preregistered Experiment 3) were  
326 collected over a period of four weeks.  
327

328 Observers viewed the stimuli with both eyes.  
329

### 330 **Observer Recruitment and Exclusion**

331  
332 Observers were recruited from the University of Pennsylvania and the local Philadelphia community and  
333 were compensated for their time. Observers were screened to have normal visual acuity (20/40 or better;  
334 with corrective eyewear, if applicable) and normal color vision, as assessed with pseudo-isochromatic  
335 plates (Ishihara, 1977). These exclusion criteria were specified in the preregistration document (see  
336 Methods: Preregistration). One observer was discontinued at this point for not meeting the normal visual  
337 acuity criterion.  
338

339 Observers who passed the vision screening then participated in a practice session. This session also served  
340 to screen for observers' ability to reliably perform the psychophysical task. At the beginning of the  
341 practice session, observers were familiarized with the task via a familiarization block. In the  
342 familiarization block, observers performed 40 trials of the task using images with covariance scalar 0.00  
343 (10 easy trials, 10 moderate trials, and 20 regular trials). In the easy trials, the observers compared images  
344 with target object LRF 0.35 and 0.45. In the moderate trials, they compared images with target object  
345 LRF 0.40 to images with target object LRF 0.35 or 0.45. In the regular trials they compared images with  
346 target object LRF 0.40 to images with target object LRF in the range [0.35, 0.45]. The data from the  
347 familiarization block was not saved. The observer then performed three normal blocks for images with  
348 covariance scalar 0.00. At the end of the practice session, the mean threshold of the observer for the last  
349 two blocks was computed. The observer was excluded from further participation if their mean threshold  
350 for the last two blocks in the practice session exceeded 0.025 ( $\log T^2$ , -3.2). This exclusion criterion was  
351 specified in our preregistered protocol (See Methods: Preregistration).  
352

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

354

## 355 Observer Information

356

357 A total of 17 observers participated in the practice sessions for the control and main experiments  
358 (preregistered Experiments 2 and 3). To de-identify observer information in the data, observers were  
359 numbered in the order they performed the practice sessions. 10 observers participated in the practice  
360 sessions for the main experiment (preregistered Experiment 3; 6 Female, 4 Male; age 18-56; mean age  
361 30.7). Four of these observers (Observer 2, Observer 4, Observer 8, and Observer 17) met the  
362 performance criterion set for screening (2 Female, 2 Male; age 23-56; mean age 38.25). All observers  
363 who advanced to the practice session had normal or corrected-to-normal vision (20/40 or better in both  
364 eyes, assessed using Snellen chart) and normal color vision (0 Ishihara plates read incorrectly). The visual  
365 acuities of the observers in the main experiment were: Observer 2, L = 20/30, R = 20/30; Observer 4, L =  
366 20/15, R = 20/20; Observer 8, L = 20/30, R = 20/25; Observer 17, L = 20/20, R = 20/20. Observers 2, 8,  
367 and 17 wore personal corrective eyewear both during vision testing and during the experiments. Observer  
368 4 did not require or use corrective eyewear.

369

## 370 Apparatus

371 The stimuli were presented on a calibrated LCD color monitor (27-in. NEC MultiSync PA271W; NEC  
372 Display Solutions) in an otherwise dark room. The monitor was driven at a pixel resolution of 1920 x  
373 1080, a refresh rate of 60Hz, and with 8-bit resolution for each RGB channel. The host computer was an  
374 Apple Macintosh with an Intel Core i7 processor. The experimental programs were written in MATLAB  
375 (MathWorks; Natick, MA) and relied on routines from the Psychophysics Toolbox  
376 (<http://psychtoolbox.org>) and mgl (<http://justingardner.net/doku.php/mgl/overview>). Responses were  
377 collected using a Logitech F310 gamepad controller.

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

## 381 Monitor Calibration

382 The monitor was calibrated using a spectroradiometer (PhotoResearch PR650). To calibrate the monitor,  
383 we focused the spectroradiometer on a patch displayed on the center of the monitor. The patch size was  
384 4.66cm x 4.66cm (3.56° x 3.56°). The optics of the radiometer sampled the emitted light from a 1°  
385 circular spot within the patch. The spectral power distribution of the three monitor primaries was  
386 measured in the range 380nm to 780nm at 4nm steps. The gamma functions for each primary were  
387 determined from measurements of the spectral power distribution for each primary at 26 equally spaced  
388 input values for that primary, in the range [0, 1] where 1 corresponds to the maximum input value of the  
389 device. These gamma functions as well as the light emitted by the monitor for an input of 0 were  
390 accounted for in the stimulus display procedures. The spectral power distribution was also measured for  
391 32 different combinations of RGB input values. These measurements were used to check the performance  
392 of the display. The maximum absolute deviation of the x-y chromaticity between the measured values and  
393 those predicted from the calibration was 0.0028 and 0.0027 for x and y chromaticity respectively, and less  
394 than 1% for luminance.

## 395 Stimulus Presentation

396

397 The size of each image was 2.6cm x 2.6cm on the monitor, corresponding to 2° by 2° visual angle. The  
398 target object size on the screen in the 2D images was ~1° in diameter. Each image was presented for

399 250ms (this was a deviation from the preregistration document, which specifies the presentation time as  
400 500ms), with an inter-stimulus interval of 250ms and inter-trial interval of 250ms. Inter-stimulus interval  
401 (ISI) is defined as the interval between the first and the second image presented on each trial. The  
402 response for each trial was collected after both the images had been displayed and removed from the  
403 screen. The observer could take as long as they wished before entering the response. Feedback was  
404 provided via tones presented after the response to allow observers to maximize their performance. The  
405 next trial was presented 250ms (ITI) after the feedback. Thus, the actual inter-trial interval depended on  
406 the response time of the observer.  
407

## 408 **Psychometric Function**

409 The proportion comparison chosen data was used to obtain the psychometric function for each block.  
410 Each block consisted of 330 trials with 30 trials at each comparison lightness level. At each lightness  
411 level, we recorded the number of times the observers chose the comparison image to be lighter. The  
412 proportion comparison chosen data were fit with a cumulative normal using the Palamedes toolbox (Prins  
413 & Kingdom, 2018) to obtain four parameters of the psychometric function: threshold, slope, lapse rate  
414 and guess rate. The lapse rate was constrained to be equal to the guess rate and to be in the range [0,  
415 0.05]. The psychometric function was fit using the maximum likelihood method. The threshold was  
416 obtained as the difference between the LRFs at proportion comparison chosen 0.76 and 0.50 as obtained  
417 from the cumulative normal fit.

## 418 **Ethics Statement**

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

## 421 **Code and Data Availability**

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

## 428 **3 MODEL**

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

438 We first introduce an analytic formulation, derived in the context of signal detection theory (SDT)  
439 formulation). We then show how this can be instantiated as a linear receptive field model whose

440 performance can be simulated (LINRF formulation). An important advantage of the LINRF formulation is  
441 that it can accommodate the physical-realizability constraint incorporated into our statistical model of  
442 naturally occurring reflectances.

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

#### 454 **SDT Model Formulation**

455 We first formulate our model in the context of signal detection theory (Green, 1966). We model the visual  
456 response to the target object in each image by a univariate internal representation denoted by the variable  
457  $z$ . This variable depends on the image and is perturbed by noise. We assume that for any fixed image,  $z$  is  
458 a normally-distributed random variable whose mean depends on the target object LRF. For each image,  
459 we assume that  $z$  is perturbed on a trial-by-trial basis by independent zero-mean normally-distributed  
460 noise, and we assume that the variance of this noise is the same for the response to all images. We refer to  
461 the noise that perturbs  $z$  for a fixed image as the internal noise and denote its variance as  $\sigma_i^2$ . For each  
462 trial of the experiment,  $z$  takes on two values,  $z_s$  and  $z_c$ , one for the interval containing the standard and  
463 the other for the interval containing the comparison.

464 If we consider performance for a particular pair of target standard and comparison LRFs, performance  
465 depends both on the difference between the expected values of  $z$  for each pair of LRFs,  $\mu_s$  and  $\mu_c$ , and on  
466 the value of  $\sigma_i^2$ . In our experimental design we have ensembles of images with different backgrounds, for  
467 each value of the target object LRF and background covariance scalar. The fact that we draw  
468 stochastically from these ensembles on each trial introduces additional variability into the value of the  
469 decision variable  $z$  that corresponds to a fixed target LRF. We call this the external variability, and model  
470 it as a normal random variable with zero mean and variance  $\sigma_e^2$ . We assume that  $\sigma_e^2$  depends on the  
471 experimentally chosen covariance scalar, but not on the target sphere LRF. Thus, the distributions of  $z_s$   
472 and  $z_c$ , for a particular choice of target standard and comparison LRF and covariance scalar, are given by  
473  $P(z_s) = N(\mu_s, \sigma_t)$  and  $P(z_c) = N(\mu_c, \sigma_t)$ . Here  $\mu_s$  is the mean value of the internal representation to the  
474 standard image and  $\mu_c$  is the mean value of the internal representation to the comparison image. The  
475 overall standard deviation  $\sigma_t$  is obtained via  $\sigma_t^2 = \sigma_i^2 + \sigma_e^2$ , where  $\sigma_i^2$  and  $\sigma_e^2$  are the variance of the  
476 internal and external noise.

477 For a 2AFC discrimination task in the context of signal detection theory, the observer makes their  
478 decision based on a comparison of  $z_s$  and  $z_c$ , choosing the interval with the higher value of  $z$  as that with  
479 the higher stimulus value. The observer's sensitivity depends on the mean values and the variance of  $z$ ,  
480 and is captured by the quantity d-prime:  $d' = (\mu_c - \mu_s)/\sigma_t$ . D-prime measures the distance between the  
481 two distributions in standard deviation units. A value of  $d' = 0$  corresponds to an inability to distinguish  
482 between the standard and the comparison image. Larger values of  $d'$  indicate increasing discriminability.

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

484  $(\mu_c - \mu_s) = d' \sigma_t = d' \sqrt{(\sigma_i^2 + \sigma_e^2)} .$  (1)

485 We further assume that the difference in mean value of the internal variable ( $\mu_c - \mu_s$ ) is proportional to  
 486 the difference in the LRFs of the target object in the standard and comparison images ( $\Delta_{\text{LRF}}$ ). That is,  
 487  $(\mu_c - \mu_s) = C \Delta_{\text{LRF}}$ , where  $C$  is the proportionality constant. This yields  
 488

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

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

493 In our experiment, external variability was induced by changing the reflectance of the objects in the  
 494 background. We used a multivariate normal distribution to generate the reflectance functions of the  
 495 background objects.<sup>5</sup> To change the amount of external noise, we scaled the covariance of the  
 496 multivariate normal distribution by multiplying its covariance matrix with a scalar. Thus, for our  
 497 experiments we have

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

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

502  
 503 Converting the equation above to the form we use to represent the data, we have

504  $\log(\Delta_{\text{LRF}}^2) = \log(\sigma_i^2 + \sigma^2 \times \sigma_{e0}^2) .$  (4)

505 The equation above predicts that the form of threshold  $\log(\Delta_{\text{LRF}}^2)$  as a function of covariance scalar  $\sigma^2$   
 506 should increase monotonically. For small values of  $\sigma^2$  ( $\sigma^2 \ll \sigma_i^2/\sigma_{e0}^2$ ), the threshold will approach a  
 507 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  
 508 approach a straight line with slope 1 in the  $\log(\Delta_{\text{LRF}}^2)$  versus  $\log(\sigma^2)$  plot. Fitting the measurements with  
 509 Equation 4 allows us to check whether the model describes the data, as well as to determine the two  
 510 parameters  $\sigma_i^2$  and  $\sigma_{e0}^2$ . In particular, we can establish the relative contribution of the internal  
 511 representational variability and external stimulus variability in limiting lightness discrimination. The  
 512 parameter  $\sigma_{e0}^2$  quantifies how much the variation in background object reflectances intrudes on the  
 513 internal representation  $z$  that mediates the lightness discrimination task. The value of  $\sigma_{e0}^2$  may be  
 514 compared directly to the intrinsic precision of that representation characterized by  $\sigma_i^2$ .

515 **Equivalent Noise Level**

516 The SDT formulation allows us to introduce the concepts of equivalent noise and equivalent noise level.  
 517 The equivalent noise is the amount of external variation that has the same effect on the decision variable  $z$

---

<sup>5</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 in the LINRF formulation below.

518 as the internal noise. The external variation is characterized experimentally by the covariance scalar  
 519 (together with the underlying model of natural reflectances which is held fixed across the experiments).  
 520 Once the model parameters  $\sigma_i^2$  and  $\sigma_{e0}^2$  are determined from the data, we can find the covariance scalar  
 521  $\sigma_{equiv}^2$  that produces externally-generated equivalent noise

522 
$$\sigma_{equiv}^2 = \sigma_i^2 / \sigma_{e0}^2. \quad (5)$$

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

525 
$$\sigma_{enl}^2 = \sigma^2 / \sigma_{equiv}^2. \quad (6)$$

526 For  $\sigma_{enl}^2 \ll 1$ , the effect of the external noise is negligible and does not affect the perceptual  
 527 representation and the internal noise dominates the precision of the representation. For  $\sigma_{enl}^2 \gg 1$ , the  
 528 effect of the external noise dominates the perceptual representation, and the visual system has not  
 529 insulated the representation of the task-relevant stimulus variable from the variation in the task-irrelevant  
 530 perceptual variable. When the equivalent noise level is  $\sim 1$ , the effect of the external variability is matched  
 531 to that of the internal variability. At this operating point, further insulation of the task-relevant  
 532 representation will not lead to significant further increases in the precision of this representation. We can  
 533 thus use the equivalent noise level as a calibrated metric for assessing the magnitude of the perceptual  
 534 effect of various levels of task-irrelevant stimulus variation.

### 535 Linear Receptive Field Formulation

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

545 We begin with development that connects the LINRF formulation to the SDT formulation. In the LINRF  
 546 formulation, the decision variable is computed from the displayed stimulus as the response of a single unit  
 547 whose responses are a linear function of the stimulus image. Denote the stimulus image by the column  
 548 vector  $I$ , and the receptive field by the column vector  $R$ . The entries of  $I$  are the radiant power emitted by  
 549 the monitor at each image location. The entries of  $R$  are the corresponding sensitivities of the linear  
 550 receptive field to each entry of  $I$ . The response of the receptive field is given as  $r_i = R^T I + \eta_i$ , where  $\eta_i$   
 551 is a random variable representing a draw of zero mean normally-distributed internal noise (variance  $\sigma_{ri}^2$ )  
 552 in the receptive field response for a fixed image. We assume that  $\sigma_{ri}^2$  is independent of  $I$ .

553 Denote  $I_{s0}$  and  $I_{c0}$  as the standard and comparison images without external noise. External normally-  
 554 distributed noise is added to both  $I_{s0}$  and  $I_{c0}$ , with covariance matrix  $\Sigma_e$ . The external noise need not have  
 555 zero mean. After incorporation of the external noise, the response of the receptive field to the comparison  
 556 and standard images is given by

557 
$$r_{ic} = R^T (I_{c0} + \eta_e) + \eta_i = R^T I_{c0} + \eta \quad (7)$$

$$558 \quad r_{is} = R^T(I_{so} + \eta_e) + \eta_i = R^T I_{so} + \eta. \quad (8)$$

559 Here  $\eta_e$  is a random variable representing a draw of external noise,  $\eta_i$  represents the internal noise, and  $\eta$   
 560 is a random variable representing the overall effect of the external and internal noise. Since the receptive  
 561 field and noise models are linear and normal,  $\eta$  is normal with variance

$$562 \quad \sigma_{\eta}^2 = (\sigma_{ri}^2 + R^T \Sigma_e R). \quad (9)$$

The mean difference between the receptive field response to the comparison and the standard image is given by  $(\mu_c - \mu_s) = R^T(I_{c0} - I_{s0}) = C'\Delta_{LRF}$ . Here  $I_{s0}$  and  $I_{c0}$  are the standard and comparison images without external noise added,  $C'$  is a constant, and  $\Delta_{LRF}$  is as defined in the SDT section above. The second equality follows because 1) the difference between  $I_{c0}$  and  $I_{s0}$  is proportional to  $\Delta_{LRF}$  as only the target LRF changes between these two images and 2) even if the mean of the external noise is non-zero, its effect cancels when we obtain the mean difference in response.

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

$$573 \quad \Delta_{\text{LRF}} = \frac{d'}{C'} \sqrt{\sigma_{ri}^2 + \sigma^2 \times (R^T \Sigma_{e0} R)} \quad (10)$$

where we have introduced the covariance scalar  $\sigma^2$  in the term corresponding to the variance of the external noise, and where  $\Sigma_{e0}$  denotes the covariance matrix of the external noise corresponding to the level of variation in natural images. Comparing to relation derived in the SDT model (Equation 3), we see that this is the same functional form for the relation between  $\Delta_{\text{LRF}}$  and  $\sigma^2$  as derived there, where we associate  $\sigma_i^2 = \frac{\sigma_{ri}^2}{(C')^2}$  and  $\sigma_{e0}^2 = \frac{(R^T \Sigma_{e0} R)}{(C')^2}$ .

To fit the LINRF formulation and relax its assumptions, we compute how images produce retinal cone excitations and employ a one-parameter description of a simple center-surround receptive field that draws upon the output of the cones. We use simulation to compute model responses for any choice of  $\sigma_i^2$ . This procedure is described in more detail below. Once the fitting procedure establishes  $R$  and  $\sigma_i^2$  that best account for the data, we then find  $\sigma_{e0}^2$  directly by passing the images corresponding to  $\sigma^2 = 1$  through the receptive field and finding the resulting variance. These parameters in turn allow us to compute  $\sigma_{equiv}^2$  and  $\sigma_{enl}^2$  for the LINRF formulation.

586 Fitting the SDT Model Formulation

The model was fit to the threshold versus covariance scalar data to obtain the parameters  $\sigma_i^2$  and  $\sigma_e^2$ . The parameters were obtained by minimizing the mean squared error between the measured and predicted threshold using the MATLAB function *fmincon*. The best fitting parameters were estimated separately for the mean observer and the individual observers.

591 Fitting the Linear Receptive Field Model Formulation

592 We fit the linear receptive field (LINRF) model using a simulation approach. We used simulation for two  
 593 reasons. First, it allows us to incorporate a model of the early visual system into the computations.

594 Second, it provides a way to account for truncation in the normally-distributed model of natural  
595 reflectances.

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

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

620 The threshold predictions of the LINRF formulation for any choice of model parameters were obtained  
621 using simulation of a two-interval forced choice paradigm similar to the experiment. For each trial, we  
622 randomly sampled a standard image and a comparison image from our dataset, following the procedure  
623 used in the experiment. We obtained the response of the receptive field (noise-added dot product) to the  
624 images and compared them to determine the simulated choice on that trial. This process was repeated  
625 10,000 times for each of the 11 comparison LRF levels. The proportion comparison chosen data were  
626 used to fit the psychometric function and obtain the discrimination threshold, similar to the method used  
627 for the human psychophysical data. We estimated model threshold for the six values of covariance scalar  
628 at which we performed the human experiments.

629 We calculated the mean squared error (averaged over the six covariance scalar values) between the  
630 thresholds of the human data being fit and the computational model for a large set of values of the two  
631 model parameters: the variance of the decision noise ( $\sigma_{ri}$ ) and the value of the RF surround ( $v_s$ ). The  
632 mean squared error values obtained as a function of these two parameters were fit with a degree two  
633 polynomial of two variables using the MATLAB *fit* function. The resulting polynomial was evaluated to  
634 estimate the parameters with lowest mean square error. These parameters were then used to estimate the  
635 internal and external noise standard deviation of the LINRF formulation using the relations:  $\sigma_i^2 = \frac{\sigma_{ri}^2}{(C')^2}$   
636 and  $\sigma_{e0}^2 = \frac{(R^T \Sigma_{e0} R)}{(C')^2}$  as explained above, where the constant  $C'$  was obtained by solving  $R^T(I_{c0} - I_{s0}) =$   
637  $C' \Delta_{\text{LRF}}$ .

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

639 **4 RESULTS**

640

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

642

643 We measured lightness discrimination thresholds of human observers as a function of the amount of  
644 variation in the reflectance spectra of the background objects in the scene. The amount of variation was  
645 determined by the covariance matrix of the multivariate normal distribution from which the spectra were  
646 sampled. We controlled the variance by multiplying the covariance matrix by a covariance scalar ( $\sigma^2$ ).  
647 We measured discrimination thresholds of four human observers at six values of the covariance scalar.  
648 The threshold was measured three times (three separate blocks) for each observer and for each value of  
649 covariance scalar. The psychometric functions for each block/covariance scalar value are shown for one  
650 observer in Figure 4 and for all observers in Figure S3. Inspection of the psychometric functions shows  
651 that their slopes steadily decrease with increasing covariance scalar, corresponding to an increase in  
652 thresholds.

653

654 Figures 5 and 6 show the data in more digested form. These plots show explicitly how the discrimination  
655 thresholds change with the amount of variability in the reflectance of the background objects. In Figure 5,  
656 mean log threshold squared (averaged across observers,  $N = 4$ ) is plotted against the log of the covariance  
657 scalar. Figure 6 plots thresholds in the same format for the individual observers, with the data averaged  
658 over the three blocks for each covariance scalar. The choice to plot the data as log threshold-squared  
659 against the log of the covariance scalar was motivated by the relatively simple expression of the SDT  
660 model formulation's predictions for this representation (see Equation 4 and following text). Table S2  
661 provides the thresholds and SEMs from Figure 6 in tabular form.

662

663 For low values of the covariance scalar, the thresholds are nearly constant and are similar across  
664 observers. As the covariance scalar increases, log squared threshold rises. These features are seen in the  
665 mean data (Figure 5) and in the data for all observers (Figure 6). The covariance scalar value at which  
666 thresholds start to increase is also similar across observers. There is some individual variability, however,  
667 in the slope of the rising limb of the measured functions.

668

669 **Modeling the Impact of Background Reflectance Variation**

670

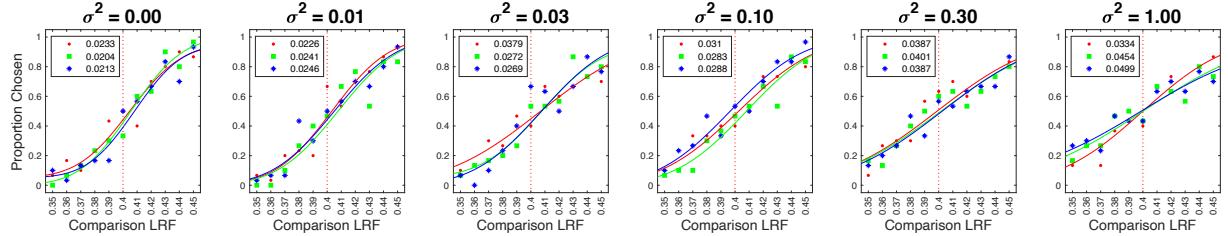
671 To interpret the data further, we fit the data with two formulations of our model (see Model section  
672 above). The performance of both the SDT and LINRF model formulations is determined by two  
673 fundamental factors. The first factor is variability in the perceptual representation of lightness internal to  
674 the visual system (i.e., internal noise, model parameter  $\sigma_i^2$ ). The second factor is the effect of  
675 experimentally-induced task-irrelevant stimulus variability (i.e., background object reflectance  
676 variability) on the same perceptual representation (i.e. external noise, model parameter  $\sigma_{e0}^2$ ). Roughly  
677 speaking, threshold with no external variation (covariance scalar  $\sigma^2 = 0$ ) establishes the level of the  
678 internal noise, while the way threshold increases with covariance scalar determines  $\sigma_{e0}^2$ . The fits  
679 determine the parameters of the model as well as allow us to examine how well the model fits the data.

680

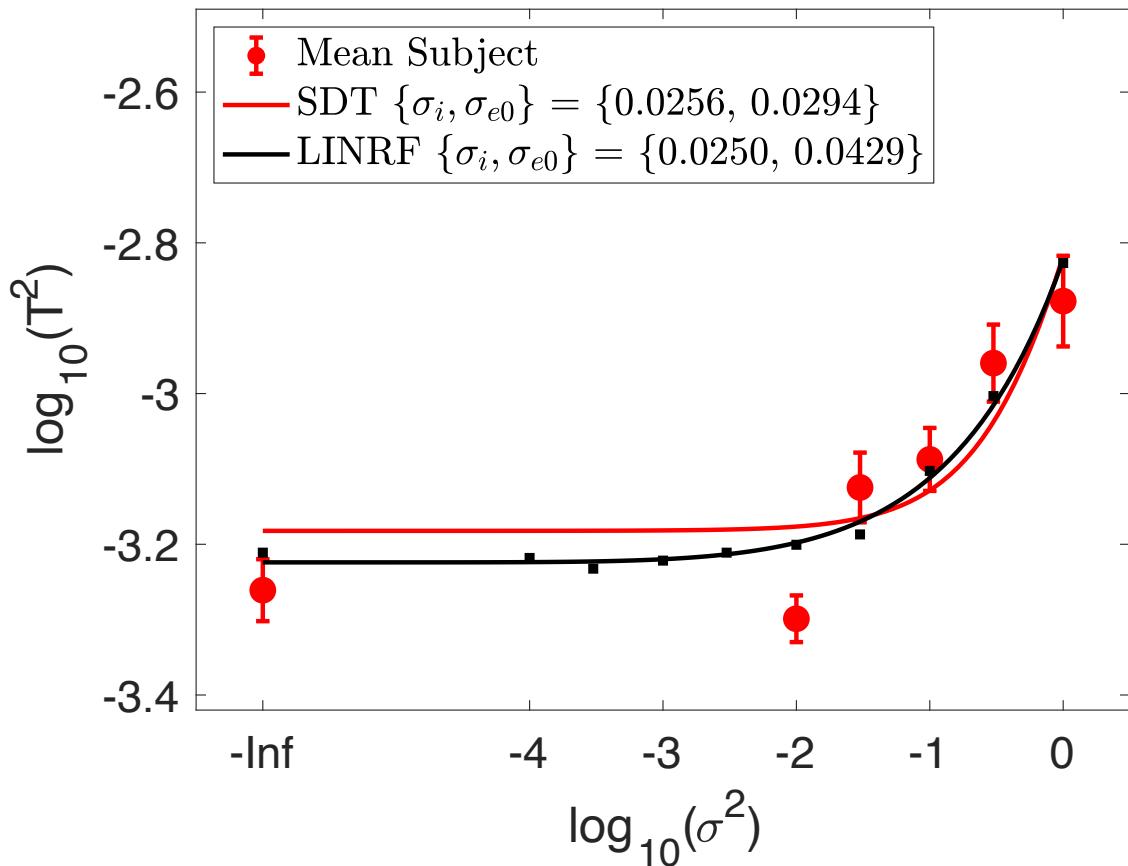
681 The fits to the mean observer data are shown in Figure 5; the fits to the individual observer data are  
682 shown in Figure 6.

683

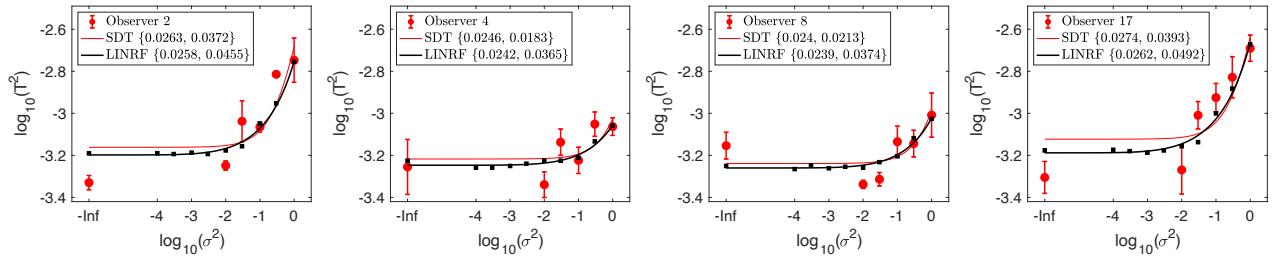
684 The fit of the analytic STD formulation (red curves) captures the main trends in the mean data and  
685 similarly for the fits to the individual observer data. Detailed examination, however, reveals that this  
686 formulation tends to overestimate thresholds in the low covariance scalar regime. An alternative way of



**Figure 4: Psychometric functions for Observer 2.** We measured the proportion comparison chosen data at six values of the covariance scalar ( $\sigma^2$ ), separately in three blocks for each observer. The data for each block was fit with a cumulative normal to obtain the discrimination threshold (see Figure 2). Each panel plots the measured values and the cumulative fit to the proportion comparison data for each of the three blocks, for Observer 2. The values in the legend provide the estimate of lightness discrimination threshold for each block obtained from the cumulative fit. See Figure S3 for the psychometric functions of all observers.



**Figure 5: Background variation increases lightness discrimination threshold.** Mean ( $N = 4$ )  $\log$  squared threshold vs  $\log$  covariance scalar from the human psychophysics (red circles). The error bars represent  $\pm 1$  SEM taken between observers. The fit of the STD formulation of the model (Equation 4) is shown as the red curve. The parameters corresponding to this fit are provided in the legend. The threshold of the fit linear receptive field (LINRF) formulation was estimated by simulation at 10 logarithmically spaced values of the covariance scalar (black squares). The black smooth curve is a smooth fit to these points of the functional form  $\log_{10} T^2 = a + b^{(x+c)^d}$  where  $x = \log_{10} \sigma^2$  and  $a, b, c$  and  $d$  are parameters adjusted in the fit. This functional form was chosen simply to provide a smooth curve through the simulated thresholds and has no theoretical significance. The parameters of the LINRF fit are also provided in the legend.



**Figure 6: Threshold of individual human observers.** Mean (across sessions) squared threshold vs log covariance scalar for individual human observers. Same format as Figure 5; here the error bars represent  $\pm 1$  SEM taken across the three blocks for each observer. The parameters of the SDT and LINRF formulations were obtained separately for each observer and are provided in the legend, in order  $\sigma_i^2, \sigma_{e0}^2$ .

687 putting this is that it underestimates the rising slopes as covariance scalar increases. Because the rising  
688 slope of this formulation asymptotes to 1, the SDT formulation of the model is not able to simultaneously  
689 describe thresholds over the full covariance scalar range.

690  
691 The fits of the LINRF formulation (black curves) are better. The fit to the mean data does an excellent job  
692 of capturing these data, and the fits to the individual observer data are also improved relative to the SDT  
693 formulation. We attribute the improvement in fit of the LINRF formulation primarily to its ability to  
694 account for the truncation of our experimental reflectance distributions, which the SDT formulation  
695 cannot do (see Section 3 Model).

696  
697 The model fits provide estimates of internal and external noise for the human observers in this task.  
698 Figure 7 (left panel) plots the estimates of the internal and external noise standard deviations (quantities  
699  $\sigma_i$  and  $\sigma_{e0}$ , for both the SDT model and the LINRF formulation. There is good consistency in the value of  
700  $\sigma_i$  across observers, the model's manifestation of the observations that thresholds for low covariance  
701 scalars are similar across observers. There is more variability in  $\sigma_{e0}$  across observers, corresponding to  
702 the individual variability seen in the rising limb of the threshold versus covariance scalar plots.

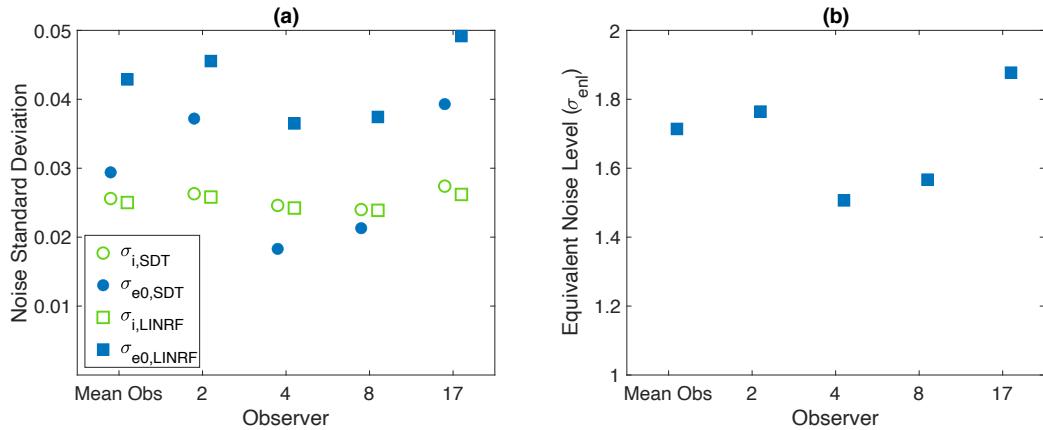
703  
704 The Poisson noise included in the LINRF formulation does not typically limit human discrimination  
705 performance at daylight light levels (Banks, Geisler, & Bennett, 1987; Cottaris, Jiang, Ding, Wandell, &  
706 Brainard, 2019). Thus, it is not surprising that the mean values of the internal noise standard deviation  
707 parameter  $\sigma_i$  for the LINRF formulation are close to those obtained with the SDT formulation (SDT  
708 formulation: mean value of internal noise standard deviation across observers 0.0256, value from fit to  
709 mean data 0.0256; LINRF formulation: mean value of internal noise standard deviation across individual  
710 observers 0.0250, value from fit to mean data, 0.0250).

711  
712 The estimates of the external noise standard deviation parameter  $\sigma_{e0}$  are higher for the LINRF  
713 formulation than for the SDT formulation (SDT formulation: mean value of external noise standard  
714 deviation 0.0290, value from fit to mean data across observers 0.0294; LINRF formulation: mean value of  
715 external noise standard deviation across observers 0.0421, value from fit to mean data, 0.0429). This is  
716 consistent with the observation that the SDT formulation underestimates the rise in thresholds with  
717 increasing covariance scalar, while this rise is captured more accurately by the LINRF formulation,  
718 presumably because the latter incorporates the constraint that the reflectance values at each wavelength  
719 are physically realizable (i.e., reflectances lie between 0 and 1).

720  
721 If we focus on the estimates from the better fitting LINRF formulation, we can compute the equivalent  
722 noise level ( $\sigma_{enl}$ ) corresponding to covariance scalar  $\sigma^2 = 1$ , the level of background object reflectance  
723 variation corresponding to our full model of natural reflectance. For the fits to the mean data, this  
724 equivalent noise level is  $\sim 1.7$ . This as well as values for the individual observers are plotted in the right  
725 panel of Figure 7. This tells us that, for our experimental conditions, the variability in the human  
726 representations of lightness induced by naturally occurring variation in background object reflectances is  
727 within a factor of two of the limits imposed by the intrinsic precision of that representation. Had the value  
728 been closer to 1, we would have concluded that the visual system had discounted the effect of variation in  
729 the background object reflectances about as required, given the intrinsic precision of the lightness  
730 representation. The fact the equivalent noise level is higher than 1 but not tremendously so is consistent  
731 with the idea that the visual system has a degree of lightness constancy, but that this constancy can be  
732 incomplete (see e.g., Gilchrist, 2006; Kingdom, 2011; Murray, 2021).

733  
734 **5 DISCUSSION**

735  
736 The perceived lightness of an object can depend on the scene in which it lies. Stabilization of the lightness  
representation against variation in scene properties extrinsic to the object's surface reflectance is referred



**Figure 7. Equivalent noise analysis.** **(a)** The left panel shows the parameter estimates for the two model formulations for the mean data and each individual observer. From these, we can estimate the equivalent noise level ( $\sigma_{enl}$ ) for background object reflectance variation corresponding to the full model of natural reflectance variation (covariance scalar  $\sigma^2 = 1$ ). **(b)** The equivalent noise level is provided for the mean data and each individual observer in the right panel.

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

## Relation to Contrast Detection in Contrast Noise

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

## Spatial and Chromatic Properties of the Stimuli

We used small image patches in our study. The small size of the image patches is a notable difference between our stimuli and natural viewing. In this initial deployment of our paradigm, we thus focused on effects of background object reflectance variation that are nearby the test object. The observed effects may be mediated by relatively small populations of neurons. The use of small image patches is not a

788 necessary requirement of our paradigm, which could be extended to larger images. Such extension could  
789 reveal additional effects not captured by the current experiments.

790  
791 In addition to using small patches, we did not vary the spatial structure of the array of objects in the  
792 rendered scenes. Manipulating spatial structure, in addition to increasing image size, may provide a way  
793 to use our paradigm to measure the spatial tuning of the mechanism(s) mediating the background effect.  
794 This approach is loosely analogous to how manipulating the structure of contrast noise may be used to  
795 examine the tuning of mechanisms supporting the detection of contrast-defined targets (Henning, Hertz,  
796 & Hinton, 1981; Rovamo, Franssila, & Nasanen, 1992; Losada & Mullen, 1995; Nachmias, 1999;  
797 Rovamo, Raninen, & Donner, 1999).

798  
799 Although we restricted our measurements to lightness discrimination thresholds, our variation of the  
800 reflectance properties of the background objects was not limited to variation in overall reflectance. The  
801 choice to introduce background object reflectance variation along more spectral dimensions (affecting e.g.  
802 background object hue and saturation) than used for target object variation was somewhat arbitrary – we  
803 could have restricted the background object reflectance variation to one dimension (e.g. overall scale of  
804 reflectance spectra) or studied discrimination of additional (e.g. chromatic) dimensions of target object  
805 variation. As with the case of spatial structure above, extending the measurements to a wider range of  
806 stimuli is of interest. Indeed, it may be possible to manipulate the chromatic structure of the variation in  
807 background object reflectances with the goal of understanding the chromatic tuning of the background  
808 object reflectance variation’s effect on the lightness discrimination thresholds, as well as on other target  
809 object discriminations. This would again be analogous to how noise-based approaches have been used to  
810 characterize chromatic tuning of mechanisms that support the detection of chromatically-defined contrast  
811 targets (Gegenfurtner & Kiper, 1992; Sankeralli & Mullen, 1997; Julianini & Eskew, 1998; Monaci,  
812 Menegaz, Süssstrunk, & Knoblauch, 2004).

#### 813

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

#### 815

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

826  
827 Previous authors have suggested that lightness constancy improves with increasing background  
828 “articulation”. That is, increasing the number of objects in the background and/or the degree to which  
829 their reflectance varies tends to improve constancy (Gilchrist, 2006; Radonjić & Gilchrist, 2013; see also  
830 Radonjić, Cottaris, & Brainard, 2015; Kraft, Maloney, & Brainard, 2002). This may on the surface seem  
831 in contradiction to our results; we find increasing the variance of the background reflectances has a  
832 deleterious effect on lightness discrimination performance. Note, however, that articulation is thought to  
833 improve constancy when the task-irrelevant variation is a change in illumination, and where the  
834 background itself is held fixed across this change. In our experiments, the illumination is held fixed and  
835 we consider the effect of the background per se, with the background change occurring across the two  
836 intervals of each forced-choice trial. Thus, we are studying a different aspect of lightness constancy than  
837 where increased articulation is thought to lead to improvements, and our results are not in conflict with  
838 previous findings.

839  
840 Our paradigm could be used to study constancy across changes in illumination, if the task-irrelevant  
841 variation used in the experiment were in the illumination rather than the background object reflectances.  
842 In that case, the articulation idea would predict a smaller elevation of lightness discrimination thresholds  
843 when the effect of illumination variation was studied for scenes with higher variance in the background  
844 reflectance, as long as the background was held fixed across the two intervals of each trial.  
845

## 846 Applications to Understanding Neural Mechanisms

847

848 A longstanding goal of vision science is to connect psychophysical performance to its underlying neural  
849 mechanisms. For probing mechanisms that mediate perceptual constancies, our paradigm has the  
850 attractive feature that there is a well-defined correct answer on each trial, so that for studies with animal  
851 subjects it is possible to provide performance-contingent reward. In addition, there are well-worked out  
852 methods for predicting psychophysical discrimination performance from recordings of the responses of  
853 neural populations (Shadlen, Britten, Newsome, & Movshon, 1996; Parker & Newsome, 1998; Cohen &  
854 Newsome, 2009; Nienborg, Cohen, & Cumming, 2012; Ruff, Ni, & Cohen, 2018), and the theoretical  
855 links between such analysis and performance should continue to hold when task-irrelevant stimulus  
856 variation is added to the paradigm. Complementing neural measurements that include random,  
857 unpredictable task-irrelevant stimulus variation with such analyses may provide rigorous quantitative  
858 insights about the sensory-perceptual processing and the neural computations underlying color and  
859 lightness constancy specifically, and perceptual constancy more generally.  
860

## 861 Model of Natural Surface Reflectances

862

863 We used a truncated multivariate normal distribution as the statistical model for the projection weights of  
864 a linear model of naturally occurring reflectances, to sample the background object reflectance functions.  
865 This model was developed in our earlier work and is evaluated more fully there (Singh, Cottaris, Heasly,  
866 Brainard, & Burge, 2018; see also Brainard & Freeman, 1997; Zhang & Brainard, 2004). The model is  
867 based on measurements of the surface reflectance functions of the Munsell papers (Kelly, Gibson, &  
868 Nickerson, 1943) as well as natural surfaces characterized by Vrhel (1994). The underlying multivariate  
869 normal provides a convenient way to capture two basic aspects of natural variation in reflectance. First,  
870 these reflectances are well-described by low-dimensional linear models (Cohen, 1964; Maloney, 1986;  
871 Parkkinen, Hallikainen, & Jaaskelainen, 1989). Second, within the reflectance subspace defined by the  
872 linear models, not all reflectances are equally likely to occur. Still, we think it likely that future work will  
873 lead to more accurate statistical models of naturally occurring reflectance. For example, it is possible that  
874 replacing the linear model approach with a prior that favors spectrally-smooth reflectance functions  
875 (Jiang, Farrell, & Wandell, 2016) would lead to a more accurate characterization. In addition, we have  
876 assumed that the distribution of reflectance functions over objects is independent, but this assumption  
877 may not be accurate. Approaches to modeling a dependency have been suggested (Shen & Yeo, 2011;  
878 Gehler, Rother, Kiefel, Zhang, & Schölkopf, 2011; Barron & Malik, 2012b; Barron & Malik, 2012a).  
879

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

890

891 **Rule of Combination**

892

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

903

904 **6 ACKNOWLEDGEMENTS:** NSF BCS 2054900 (VS), NIH R01 EY10016 (DHB), NIH R01  
905 EY028571 (JB).

906

907 APPENDIX

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

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

918 The basic methods were the same as for preregistered Experiment 3. The practice session was conducted  
919 with the images in Condition 1 described below. The observers were retained for the experiment if their  
920 average threshold of the last two blocks during the practice session was lower than 0.030. This was a  
921 deviation from the preregistered plan where we set the threshold criterion as 0.025. After collecting data  
922 from 8 observers, we realized that the criterion was too strict. Only one observer had met the criterion.  
923 After modifying the threshold criterion, we included two of the initially discontinued observers in our  
924 experiment (Observer 5 and Observer 8). A total of 11 naïve observers participated in the practice  
925 sessions. Four of these observers met the criteria for continuing the experiment. Two of these observers  
926 also participated in the main experiment (Observer 4 and Observer 8). The visual acuities of these 4  
927 observers were: Observer 4, L = 20/15, R = 20/20; Observer 5, L = 20/20, R = 20/40; Observer 8, L =  
928 20/30, R = 20/25; Observer 11, L = 20/25, R = 20/30. Observers 5, 8, and 11 wore personal corrective  
929 eyewear both during vision testing and during the experiments. Observer 4 did not require or use  
930 corrective eyewear.

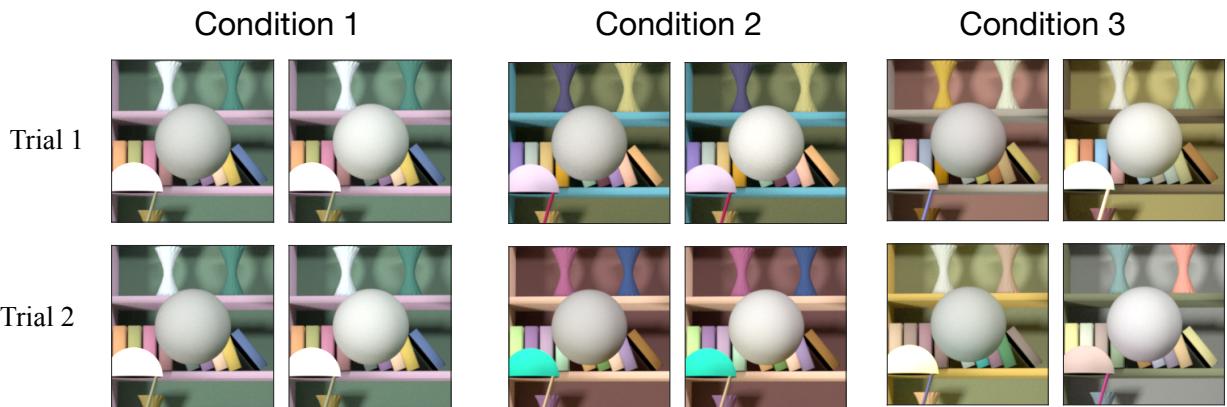
931 We measured lightness discrimination threshold of four naïve human observers using a two-interval  
932 forced choice paradigm. The thresholds were measured for three specific types of background variation  
933 (Figure S1). The reflectance spectra of the background objects were generated with the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

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

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

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

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



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

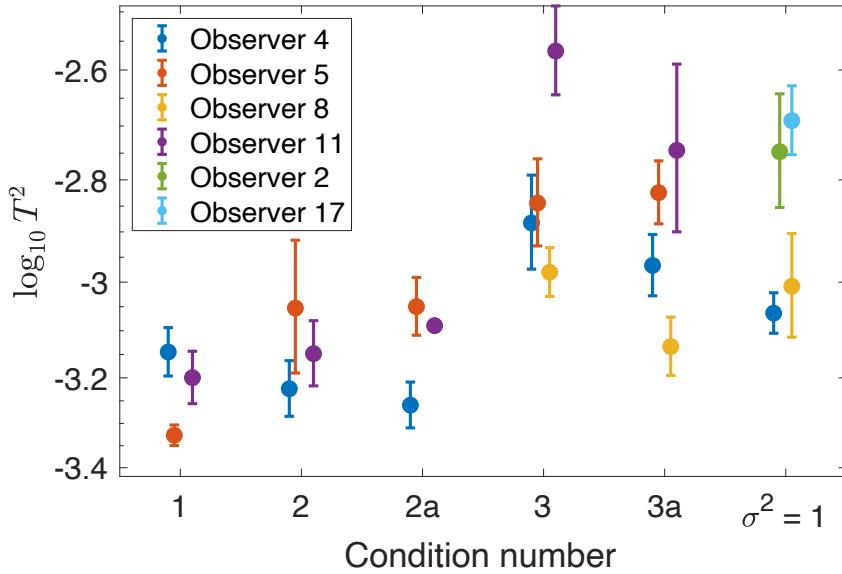
946 Condition 2a. Between-trial background variation without secondary reflection: Same as Condition 2, but  
947 without multiple reflections of light from object surfaces. The light rays only bounce off once from the  
948 surfaces before coming to the camera.

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

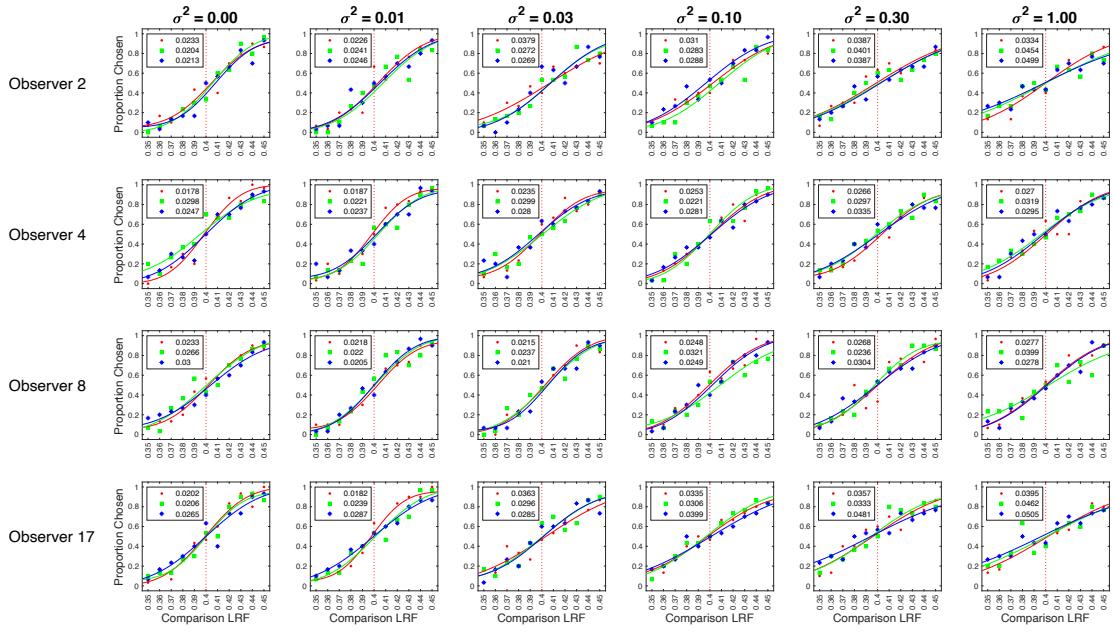
952 Figure S2 shows the discrimination thresholds of the four human observers for the five conditions studied  
953 in this experiment. We plot the mean threshold and the standard error of the mean (SEM) taken over the  
954 three separate threshold measurements. For each observer, the thresholds for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

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

973



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



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

974  
975  
976

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

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

977  
978  
979  
980  
981  
982

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

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

983  
984

985      **Figure Captions**  
986

987      **Figure 1: Psychophysical task.** **(a)** On every trial of the experiment, human observers viewed  
988      two images in sequence, a standard image and a comparison image and indicated the one in  
989      which the spherical target object in the center of the image was lighter. Example standard and  
990      comparison images are shown. The images were computer graphics simulations. The simulated  
991      reflectance functions of the target were spectrally flat, and the spheres appeared gray. The overall  
992      reflectance of the target was held fixed in the standard images and differed between standard and  
993      comparison. Performance (proportion correct) was measured as a function of this difference to  
994      determine discrimination threshold. The reflectance spectra of objects in the background could  
995      be held fixed or vary between standard and comparison on each trial (as illustrated here). The  
996      order of presentation of the standard and comparison images was randomized from trial to trial.  
997      Discrimination thresholds were measured as function of the amount of variation in background  
998      object reflectances. **(b)** Trial sequence.  $R_{N-1}$  indicates the time of the observer's response for the  
999       $(N-1)^{\text{th}}$  trial. The  $N^{\text{th}}$  trial begins 250ms after that response (Inter Trial Interval, ITI). The  $N^{\text{th}}$  trial  
1000      consists of two 250ms stimulus presentation intervals with a 250ms inter-stimulus interval (ISI).  
1001      The observer responds by pressing a button on a gamepad after the second stimulus has been  
1002      shown. The observer can take as long as he or she wishes before making the response, with an  
1003      example response time denoted by  $R_N$  in the figure. The next trial begins 250ms after the  
1004      response.  
1005

1006      **Figure 2: Psychometric function.** We recorded the proportion of times the observer chose the target in  
1007      the comparison image to be lighter, as a function of the comparison LRF. The LRF of the target object in  
1008      the standard image was fixed at 0.4. The LRF of the target object in the comparison image were chosen  
1009      from 11 linearly spaced values in the range [0.35, 0.45]. In each block, thirty trials were presented at each  
1010      comparison LRF value. We fit a cumulative normal distribution to the proportion comparison chosen data  
1011      using maximum likelihood methods. The guess and lapse rates were constrained to be equal and were  
1012      restricted to be in the range [0, 0.05]. The threshold was measured as the difference between the LRF at  
1013      proportion comparison chosen equal to 0.76 and 0.5, as predicted by the cumulative normal fit. This  
1014      figure shows the data for Observer 2 for scale factor 0.00, for the block run in the first experimental  
1015      session for that observer. The point of subjective equality (PSE, the LRF corresponding to proportion  
1016      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.

1017      **Figure 3: Variation in background object reflectances:** The reflectance spectra of background objects  
1018      were chosen from a multivariate normal distribution that modeled the statistics of natural reflectance  
1019      spectra. The variation in the reflectance spectra was controlled by multiplying the covariance matrix of  
1020      the distribution with a scalar. We generated images at six levels of the scalar. Each column shows three  
1021      sample images at each of the six values of the scalar. The leftmost column corresponds to no variation  
1022      and the rightmost column corresponds to the modeled variation of natural reflectances. The target object  
1023      (sphere at the center of each panel) in each image has the same LRF. For each value of the scalar, we  
1024      generated 1100 images, 100 each at 11 linearly spaced target LRF levels across the range [0.35, 0.45].  
1025      Discrimination thresholds were measured separately for each value of the covariance scalar.  
1026

1027      **Figure 4: Psychometric functions for Observer 2.** We measured the proportion comparison chosen data  
1028      at six values of the covariance scalar ( $\sigma^2$ ), separately in three blocks for each observer. The data for each  
1029      block was fit with a cumulative normal to obtain the discrimination threshold (see Figure 2). Each panel  
1030      plots the measured values and the cumulative fit to the proportion comparison data for each of the three  
1031      blocks, for Observer 2. The values in the legend provide the estimate of lightness discrimination threshold

1032 for each block obtained from the cumulative fit. See Figure S3 for the psychometric functions of all  
1033 observers.

1034

1035 **Figure 5: Background variation increases lightness discrimination threshold.** Mean ( $N = 4$ ) log  
1036 squared threshold vs log covariance scalar from the human psychophysics (red circles). The error bars  
1037 represent  $\pm 1$  SEM taken between observers. The fit of the STD formulation of the model (Equation 4)  
1038 is shown as the red curve. The parameters corresponding to this fit are provided in the legend. The  
1039 threshold of the fit linear receptive field (LINRF) formulation was estimated by simulation at 10  
1040 logarithmically spaced values of the covariance scalar (black squares). The black smooth curve is a  
1041 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  $a, b, c$   
1042 and  $d$  are parameters adjusted in the fit. This functional form was chosen simply to provide a smooth  
1043 curve through the simulated thresholds and has no theoretical significance. The parameters of the LINRF  
1044 fit are also provided in the legend.

1045

1046 **Figure 6: Threshold of individual human observers.** Mean (across sessions) squared threshold vs log  
1047 covariance scalar for individual human observers. Same format as Figure 5; here the error bars represent  
1048  $\pm 1$  SEM taken across the three blocks for each observer. The parameters of the SDT and LINRF  
formulations were obtained separately for each observer and are provided in the legend, in order  $\sigma_i^2, \sigma_{e0}^2$ .

1049

1050 **Figure 7. Equivalent noise analysis.** (a) The left panel shows the parameter estimates for the two model  
1051 formulations for the mean data and each individual observer. From these, we can estimate the equivalent  
1052 noise level ( $\sigma_{eni}$ ) for background object reflectance variation corresponding to the full model of natural  
1053 reflectance variation (covariance scalar  $\sigma^2 = 1$ ). (b) The equivalent noise level is provided for the mean  
data and each individual observer in the right panel.

1054

1055 **Figure S1: Control experiment stimuli.** Example stimuli for Conditions 1, 2 and 3 in the control  
1056 experiment (preregistered Experiment 2) to study the effect of variation in background object reflectances  
1057 on lightness discrimination threshold. In condition 1, the background was fixed in every trial and every  
1058 interval. In Condition 2, the background object reflectances varied from trial to trial, but remained fixed  
1059 in the two intervals of a trial. In Condition 3, the background object reflectances varied in each trial and  
1060 interval. For illustration, in this figure we have chosen the stimulus on the left to be the standard image  
1061 with target object at 0.4 LRF and the on the right to be comparison image with target object at 0.45 LRF.  
1062 In the experiment, the two images were presented sequentially in random order at the center of the screen.  
1063 Conditions 2a and 3a stimuli are similar to Conditions 2 and 3 respectively, but without secondary  
1064 reflections.

1065

1066 **Figure S2: Control experiment.** Lightness discrimination threshold of four human observers in the five  
1067 conditions in the control experiment (preregistered Experiment 2). The plotted points have been jittered  
1068 horizontally to avoid marker overlap. The thresholds are higher for the condition where the target objects  
1069 are compared against a change in background object reflectances (Conditions 3 and 3a) than when the  
1070 background is held fixed within each trial (Conditions 1, 2, 2a). Secondary reflections do not have any  
1071 significant effect on thresholds (Conditions 2a and 3a). Condition 3a of the control experiment is  
1072 equivalent to the condition of the main experiment (preregistered Experiment 3) with covariance scalar  
1073 equal to 1. The thresholds for this condition of the main experiment are plotted here for comparison  
( $\sigma^2 = 1$ ). Two observers from the control experiment also participated in the main experiment.

1074

1075 **Figure S3: Psychometric functions for all observers.** Same as Figure 4 for all observers retained in the  
main experiment.

1076

- 1077 REFERENCES  
1078
- 1079 Abrams, A. B., Hillis, J. M., & Brainard, D. H. (2007). The relation between color discrimination  
1080 and color constancy: when is optimal adaptation task dependent? *Neural Computation*,  
1081 19, 2610-2637.
- 1082 Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*,  
1083 262(December 24), 2042-2044.
- 1084 Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The*  
1085 *New Cognitive Neurosciences, 2nd edition* (pp. 339-351). Cambridge, MA: MIT Press.
- 1086 Afifi, M., Barron, J. T., LeGendre, C., Tsai, Y.-T., & Bleibel, F. (2021). *Cross-camera*  
1087 *convolutional color constancy*. Presented at the Proceedings of the IEEE/CVF  
1088 International Conference on Computer Vision,
- 1089 Allred, S. R., & Brainard, D. H. (2013). A Bayesian model of lightness perception that  
1090 incorporates spatial variation in the illumination. *Journal of Vision*, 13(7), 1-18.
- 1091 Alvaro, L., Linhares, J. M. M., Moreira, H., Lillo, J., & Nascimento, S. M. C. (2017). Robust  
1092 colour constancy in red-green dichromats. *PLoS ONE*, 12(6), e0180310.
- 1093 American Society for Testing and Materials. (2017). Standard test method for luminous  
1094 reflectance factor of acoustical materials by use of integrating-sphere reflectometers.  
1095 *Renovations of Center for Historic Preservation*, 98(A), E1477.
- 1096 Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal Of The Optical Society*  
1097 *Of America A*, 3(10), 1743-1751.
- 1098 Aston, S., Radonjić, A., Brainard, D. H., & Hurlbert, A. C. (2019). Illumination discrimination  
1099 for chromatically biased illuminations: implications for colour constancy. *Journal of*  
1100 *Vision*, 19(30:15)
- 1101 Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility.  
1102 *Vision Research*, 27(11), 1915-1924.
- 1103 Barron, J. T., & Malik, J. (2012a). *Color constancy, intrinsic images, and shape estimation*.  
1104 Paper presented at ECCV.
- 1105 Barron, J. T., & Malik, J. (2012b). *Shape, albedo, and illumination from a single image of an*  
1106 *unknown object*. Paper presented at IEEE Conference on Computer Vision and Pattern  
1107 Recognition, 334-341.
- 1108 Bloj, M., Ripamonti, C., Mitha, K., Greenwald, S., Hauck, R., & Brainard, D. H. (2004). An  
1109 equivalent illuminant model for the effect of surface slant on perceived lightness. *Journal*  
1110 *of Vision*, 4(9), 735-746.
- 1111 Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on  
1112 perceived surface albedo in binocularly viewed scenes. *Journal of Vision*, 3(8), 541-553.
- 1113 Brainard, D. H. (1989). Calibration of a computer controlled color monitor. *Color Research &*  
1114 *Application*, 14(1), 23-34.
- 1115 Brainard, D. H. (1998). Color constancy in the nearly natural image. 2. achromatic loci. *Journal*  
1116 *of the Optical Society of America A*, 15(2), 307-325.
- 1117 Brainard, D. H. (2015). Color and the cone mosaic. *Annual Review of Vision Science*, 1, 519-  
1118 546.
- 1119 Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural  
1120 image. 1. asymmetric matches. *Journal of the Optical Society of America A*, 14(9), 2091-  
1121 2110.
- 1122 Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical*  
1123 *Society of America A*, 14(7), 1393-1411.

- 1124 Brainard, D. H., Longere, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M., & Xiao, B. (2006).  
1125 Bayesian model of human color constancy. *Journal of Vision*, 6(11), 1267-1281.
- 1126 Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination  
1127 models. *Journal of Vision*, 11(5)
- 1128 Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. In J. P. Hornak  
1129 (Ed.), *Encyclopedia of Imaging Science and Technology* (pp. 172-188). New York: Wiley.
- 1130 Brainard, D. H., & Radonjić, A. (2014). Color constancy. *The New Visual Neurosciences*, 1,  
1131 545–556.
- 1132 Brascamp, J. W., & Shevell, S. K. (2021). The certainty of ambiguity in visual neural  
1133 representations. *Annual Review of Vision Science*, in press
- 1134 Brindley, G. S. (1960). *Physiology of the Retina and the Visual Pathway*. London: Arnold.
- 1135 Brown, R. O., & MacLeod, D. I. A. (1997). Color appearance depends on the variance of  
1136 surround colors. *Current Biology*, 7, 844-849.
- 1137 Burge, J. (2020). Image-computable ideal observers for tasks with natural stimuli. *Annual  
1138 Review of Neuroscience*, 6, 491-517.
- 1139 Burge, J., & Geisler, W. S. (2011). Optimal defocus estimation in individual natural images.  
1140 *Proceedings of the National Academy of Sciences*, 108(40), 16849-16854.
- 1141 Burge, J., & Geisler, W. S. (2014). Optimal disparity estimation in natural stereo images. *Journal  
1142 of Vision*, 14(2)
- 1143 Burge, J., & Geisler, W. S. (2015). Optimal speed estimation in natural image movies predicts  
1144 human performance. *Nature Communications*, 6, 7900.
- 1145 Burge, J., & Jaini, P. (2017). Accuracy maximization analysis for sensory-perceptual tasks:  
1146 computational improvements, filter robustness, and coding advantages for scaled additive  
1147 noise. *PLoS Computational Biology*, 13(2), e1005281.
- 1148 Burnham, R. W., Evans, R. M., & Newhall, S. M. (1952). Influence on color perception of  
1149 adaptation to illumination. *Journal of the Optical Society of America*, 42(9), 597-605.
- 1150 Chichilnisky, E. J., & Wandell, B. A. (1997). Increment-decrement asymmetry in adaptation.  
1151 *Vision Research*, 37, 616.
- 1152 Chin, B. M., & Burge, J. (2020). Predicting the partition of behavioral variability in speed  
1153 perception with naturalistic stimuli. *Journal of Neuroscience*, 40(4), 864-879.
- 1154 CIE. (2007). *Fundamental chromaticity diagram with physiological axes – Parts 1 and 2.*  
1155 *Technical Report 170-1*. Vienna: Central Bureau of the Commission Internationale de l'  
1156 Éclairage.
- 1157 Cohen, J. (1964). Dependency of the spectral reflectance curves of the Munsell color chips.  
1158 *Psychon. Sci.*, 1, 369-370.
- 1159 Cohen, M. R., & Newsome, W. T. (2009). Estimates of the contribution of single neurons to  
1160 perception depend on timescale and noise correlation. *J Neurosci*, 29(20), 6635-6648.
- 1161 Cottaris, N. P., Jiang, H., Ding, X., Wandell, B. A., & Brainard, D. H. (2019). A computational-  
1162 observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-  
1163 mosaic structure, and inference engine. *Journal of Vision*, 19(4), 8.
- 1164 Fechner, G. T. (1860). *Elements of Psychophysics* (H. E. Adler, 1966, Trans.). New York: Holt,  
1165 Rinehart and Winston.
- 1166 Flachot, A., & Gegenfurtner, K. R. (2018). Processing of chromatic information in a deep  
1167 convolutional neural network. *JOSA A*, 35(4), B334-B346.

- 1168 Flachot, A., & Gegenfurtner, K. R. (2021). Color for object recognition: Hue and chroma  
1169 sensitivity in the deep features of convolutional neural networks. *Vision Research*, 182,  
1170 89-100.
- 1171 Foster, D. H. (2011). Color constancy. *Vision Research*, 51(7), 674-700.
- 1172 Gegenfurtner, K., & Kiper, D. C. (1992). Contrast detection in luminance and chromatic noise.  
1173 *Journal of the Optical Society of America A*, 9(11), 1880-1888.
- 1174 Gehler, P., Rother, C., Kiefel, M., Zhang, L., & Schölkopf, B. (2011). *Recovering intrinsic*  
1175 *images with a global sparsity prior on reflectance*. Paper presented at Advances in  
1176 Neural Information Processing Systems, 765-773.
- 1177 Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual*  
1178 *Review of Psychology*, 59, 167-192.
- 1179 Geisler, W. S., Najemnik, J., & Ing, A. D. (2009). Optimal stimulus encoders for natural tasks.  
1180 *Journal of Vision*, 9(13), 17 11-16.
- 1181 Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*,  
1182 195, 185.
- 1183 Gilchrist, A. L. (2006). *Seeing Black and White*. Oxford: Oxford University Press.
- 1184 Giulianini, F., & Eskew, R. T., Jr. (1998). Chromatic masking in the (DL/L, DM/M) plane of  
1185 cone-contrast space reveals only two detection mechanisms. *Vision Research*, 38, 3913-  
1186 3926.
- 1187 Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Vol. 1). New  
1188 York: Wiley.
- 1189 Heasly, B. S., Cottaris, N. P., Lichtman, D. P., Xiao, B., & Brainard, D. H. (2014).  
1190 RenderToolbox3: MATLAB tools that facilitate physically based stimulus rendering for  
1191 vision research. *Journal of Vision*, 14(2)
- 1192 Helmholtz, H. (1896). *Physiological Optics*. New York: Dover Publications, Inc.
- 1193 Helson, H., & Jeffers, V. B. (1940). Fundamental problems in color vision. II. Hue, lightness,  
1194 and saturation of selective samples in chromatic illumination. *Journal of Experimental*  
1195 *Psychology*, 26(1), 1-27.
- 1196 Helson, H., & Michels, W. C. (1948). The effect of chromatic adaptation on achromaticity.  
1197 *Journal of the Optical Society of America*, 38, 1025-1032.
- 1198 Henning, G. B., Hertz, B. G., & Hinton, J. L. (1981). Effects of different hypothetical detection  
1199 mechanisms on the shape of spatial-frequency filters inferred from masking experiments:  
1200 I. Noise masks. *Journal of the Optical Society of America*, 71(5), 574-581.
- 1201 Hillis, J. M., & Brainard, D. H. (2005). Do common mechanisms of adaptation mediate color  
1202 discrimination and appearance? Uniform backgrounds. *Journal of the Optical Society of*  
1203 *America A*, 22(10), 2090-2106.
- 1204 Hillis, J. M., & Brainard, D. H. (2007a). Distinct mechanisms mediate visual detection and  
1205 identification. *Current Biology*, 17(19), 1714-1719.
- 1206 Hillis, J. M., & Brainard, D. H. (2007b). Do common mechanisms of adaptation mediate color  
1207 discrimination and appearance? Contrast adaptation. *Journal of the Optical Society of*  
1208 *America A*, 24(8), 2122-2133.
- 1209 Horn, B. K. P. (1974). Determining lightness from an image. *Computer Vision, Graphics, and*  
1210 *Image Processing*, 3, 277-299.
- 1211 Hurlbert, A. (2019). Challenges to color constancy in a contemporary light. *Current Opinion in*  
1212 *Behavioral Sciences*, 30:186, 186-193.
- 1213 Ishihara, S. (1977). Tests for colour-blindness. *Tokyo: Kanehara Shuppen Company, Ltd.*

- 1214 Jaini, P., & Burge, J. (2017). Linking normative models of natural tasks to descriptive models of  
1215 neural response. *Journal of Vision*, 17(12), 16.
- 1216 Jakob, W. (2010). Mitsuba Renderer.
- 1217 Jameson, D., & Hurvich, L. M. (1955). Some quantitative aspects of an opponent-colors theory.  
1218 I. Chromatic responses and spectral saturation. *Journal of the Optical Society of America*,  
1219 45, 546-552.
- 1220 Jiang, H., Farrell, J., & Wandell, B. (2016). *A spectral estimation theory for color appearance  
1221 matching*. Presented at the IS&T International Symposium on Electronic Imaging,
- 1222 Kelly, K. L., Gibson, K. S., & Nickerson, D. (1943). Tristimulus specification of the Munsell  
1223 book of color from spectrophotometric measurements. *Journal of the Optical Society of  
1224 America*, 33(7), 355-376.
- 1225 Kingdom, F. A. (2011). Lightness, brightness and transparency: a quarter century of new ideas,  
1226 captivating demonstrations and unrelenting controversy. *Vision Research*, 51(7), 652-  
1227 673.
- 1228 Knill, D. C., & Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge  
1229 University Press.
- 1230 Kraft, J. M., Maloney, S. I., & Brainard, D. H. (2002). Surface-illuminant ambiguity and color  
1231 constancy: effects of scene complexity and depth cues. *Perception*, 31, 247-263.
- 1232 Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society  
1233 of America*, 61(1), 1-11.
- 1234 Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of  
1235 the Optical Society of America A*, 4(2), 391-404.
- 1236 Losada, M. A., & Mullen, K. T. (1995). Color and luminance spatial tuning estimated by noise  
1237 masking in the absence of off-frequency looking. *Journal of the Optical Society of  
1238 America A*, 12(2), 250-260.
- 1239 Lotto, R. B., & Purves, D. (1999). The effects of color on brightness. *Nature Neuroscience*,  
1240 2(11), 1010-1014.
- 1241 Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small  
1242 numbers of parameters. *Journal Of The Optical Society Of America A*, 3(10), 1673-1683.
- 1243 Maloney, L. T., & Yang, J. N. (2001). The illuminant estimation hypothesis and surface color  
1244 perception. In R. Mausfeld & D. Heyer (Eds.), *Colour Perception: From Light to Object*  
1245 (pp. 335-358). Oxford: Oxford University Press.
- 1246 Marimont, D. H., & Wandell, B. A. (1994). Matching color images: the effects of axial  
1247 chromatic aberration. *Journal of the Optical Society of America A*, 11(12), 3113-3122.
- 1248 Monaci, G., Menegaz, G., Süssstrunk, S., & Knoblauch, K. (2004). Chromatic contrast detection  
1249 in spatial chromatic noise. *Visual Neuroscience*, 21, 291-294.
- 1250 Morimoto, T., & Smithson, H. E. (2018). Discrimination of spectral reflectance under  
1251 environmental illumination. *J Opt Soc Am A Opt Image Sci Vis*, 35(4), B244-B255.
- 1252 Murray, R. F. (2020). A model of lightness perception guided by probabilistic assumptions about  
1253 lighting and reflectance. *J Vis*, 20(7), 28.
- 1254 Murray, R. F. (2021). Lightness perception in complex scenes. *Annual Review of Vision Science*,  
1255 7, 417-436.
- 1256 Nachmias, J. (1999). How is a grating detected on a narrowband noise masker? *Vision Research*,  
1257 39(6), 1133-1142.
- 1258 Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: discrimination may be better than  
1259 detection. *Vision Research*, 14(10), 1039-1042.

- 1260 Nienborg, H., Cohen, M. R., & Cumming, B. G. (2012). Decision-related activity in sensory  
1261 neurons: correlations among neurons and with behavior. *Annu Rev Neurosci*, 35, 463-  
1262 483.
- 1263 Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. T. (2010). Categorical color  
1264 constancy for real surfaces  
. *Journal of Vision*, 10(9)
- 1265 Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of  
1266 perception. *Annual Review of Neuroscience*, 21(1), 227-277.
- 1267 Parkkinen, J. P. S., Hallikainen, J., & Jaaskelainen, T. (1989). Characteristic spectra of Munsell  
1268 colors. *Journal of the Optical Society of America*, 6(2), 318-322.
- 1269 Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. (2014). Chromatic  
1270 illumination discrimination ability reveals that human colour constancy is optimised for  
1271 blue daylight illuminations. *PLoS ONE* 9(2:e87989), e87989.
- 1272 Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding  
1273 and Efficiency* (pp. 3-24).
- 1274 Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A*,  
1275 16(3), 647-653.
- 1276 Prins, N., & Kingdom, F. A. A. (2018). Applying the model-comparison approach to test specific  
1277 research hypotheses in psychophysical research using the Palamedes toolbox. *Frontiers  
1278 in Psychology*, 9, 1250.
- 1279 Radonjić, A., & Brainard, D. H. (2016). The nature of instructional effects in color constancy. *J  
1280 Exp Psychol Hum Percept Perform*, 42(6), 847-865.
- 1281 Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015). Color constancy supports cross-  
1282 illumination color selection. *Journal of Vision*, 15(6), 1-19.
- 1283 Radonjić, A., Ding, X., Krieger, A., Aston, S., Hurlbert, A. C., & Brainard, D. H. (2018).  
1284 Illumination discrimination in the absence of a fixed surface-reflectance layout. *Journal  
1285 of Vision*, 18(5:11)
- 1286 Radonjić, A., & Gilchrist, A. L. (2013). Depth effect on lightness revisited: the role of  
1287 articulation, proximity and fields of illumination. *i-Perception*, 4(6), 437-455.
- 1288 Radonjić, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N. P., Brainard, D. H., &  
1289 Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal  
1290 of Vision*, 16(11:2), 1-18.
- 1291 Rodieck, R. W. (1998). *The First Steps in Seeing*. Sunderland, Mass.: Sinauer.
- 1292 Rovamo, J., Franssila, R., & Nasanen, R. (1992). Contrast sensitivity as a function of spatial  
1293 frequency, viewing distance and eccentricity with and without spatial noise. *Vision  
1294 Research*, 32(4), 631-637.
- 1295 Rovamo, J., Raninen, A., & Donner, K. (1999). The effects of temporal noise and retinal  
1296 luminance on foveal flicker sensitivity. *Vision Research*, 39, 533-539.
- 1297 Ruff, D. A., Ni, A. M., & Cohen, M. R. (2018). Cognition as a window into neuronal population  
1298 space. *Annual Review of Neuroscience*, 41, 77-97.
- 1299 Sankeralli, M. J., & Mullen, K. T. (1997). Postreceptor chromatic detection mechanisms  
1300 revealed by noise masking in three-dimensional cone contrast space. *Journal of the  
1301 Optical Society of America A*, 14(10), 2633-2646.
- 1302 Schultz, S., Doerschner, K., & Maloney, L. T. (2006). Color constancy and hue scaling. *J Vis*,  
1303 6(10), 1102-1116.
- 1304

- 1305 Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational  
1306 analysis of the relationship between neuronal and behavioral responses to visual motion.  
1307 *Journal of Neuroscience*, 16, 1486-1510.
- 1308 Shen, L., & Yeo, C. (2011). *Intrinsic images decomposition using a local and global sparse*  
1309 *representation of reflectance*. Presented at the IEEE Conference on Computer Vision and  
1310 Pattern Recognition,
- 1311 Singh, V., Cottaris, N. P., Heasly, B. S., Brainard, D. H., & Burge, J. (2018). Computational  
1312 luminance constancy from naturalistic images. *Journal of Vision*, 18(13), 19.
- 1313 Smithson, H. E. (2005). Sensory, computational, and cognitive components of human color  
1314 constancy. *Philosophical Transactions of the Royal Society of London. Series B*,  
1315 360(1458), 1329-1346.
- 1316 Teller, D. Y. (1984). Linking propositions. *Vision Research*, 24(10), 1233-1246.
- 1317 Thibos, L. N., Hong, X., Bradley, A., & Cheng, X. (2002). Statistical variation of aberration  
1318 structure and image quality in a normal population of healthy eyes. *Journal of the Optical  
1319 Society of America A*, 19(12), 2329-2348.
- 1320 von Kries, J. (1905). Influence of adaptation on the effects produced by luminous stimuli. In D.  
1321 L. MacAdam (Ed.), *Sources of Color Science (1970)* (pp. 120-1126). Cambridge, MA:  
1322 MIT Press.
- 1323 Vrhel, M. J., Gershon, R., & Iwan, L. S. (1994). Measurement and analysis of object reflectance  
1324 spectra. *Color Research & Application*, 19(1), 4-9.
- 1325 Wandell, B. A., & Brainard, D. H. (in press). Principles and consequences of the initial visual  
1326 encoding. In F. G. Ashby, H. Colonius & E. Dzhafarov (Eds.), *The New Handbook of  
1327 Mathematical Psychology* Cambridge: Cambridge University Press.
- 1328 Webster, M. A., & Mollon, J. D. (1995). Colour constancy influenced by contrast adaptation.  
1329 *Nature*, 373(6516), 694-698.
- 1330 Weiss, D., Witzel, C., & Gegenfurtner, K. (2017). Determinants of colour constancy and the blue  
1331 bias. *i-Perception*, 8(6), 204166951773963.
- 1332 Whittle, P., & Challands, P. D. C. (1969). The effect of background luminance on the brightness  
1333 of flashes. *Vision Research*, 9(9), 1095-1110.
- 1334 Witzel, C., & Gegenfurtner, K. R. (2018). Color perception: objects, constancy, and categories.  
1335 *Annual Review of Vision Science*, 4, 475-499.
- 1336 Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: tests of three  
1337 candidate cues. *Vision Research*, 41, 2581-2600.
- 1338 Zhang, X., & Brainard, D. H. (2004). *Bayesian color correction method for non-colorimetric*  
1339 *digital image sensors*. Paper presented at Color and Imaging Conference, 308-314.
- 1340 Zhu, H., Yuille, A., & Kersten, D. (2021). *Three-dimensional pose discrimination in natural*  
1341 *images of humans*. Presented at the Annual Meeting of the Vision Sciences Society, May  
1342 21-26, 2021. Poster A70.
- 1343