1   Title: Characterization of human lightness discrimination thresholds for independent spectral variations

2   Authors: Devin Reynolds, Vijay Singh.

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

5

6   **ABSTRACT:** Lightness of an object is an object intrinsic property that depends on its surface reflectance
7   spectrum. The human visual system infers the lightness of an object from the light reflected off its
8   surface. Since the reflected light depends also on the properties of the scene and the pose and position of
9   the observer, the visual system needs to discount any variation in the proximal signal due to these object
10   extrinsic factors. In this work, we characterize the extent to which the visual system can achieve this
11   stability for variations in the reflectance spectrum of objects and light sources in a scene. We measure
12   human observers' thresholds of discriminating images based on the lightness of a spherical target object
13   using a two alternate forced choice task. The images were computational renderings of 3D scenes. We
14   measured how the discrimination thresholds changed as we varied the reflectance spectra of the objects
15   and the intensity of the light sources in the scene, both individually and simultaneously. As the amount of
16   variability in these scene properties increased, the thresholds of discrimination first remained constant
17   indicating that the thresholds in this region depended on the variability in the observers' intrinsic
18   representation of lightness. As the extrinsic variation increased further, the thresholds started increasing
19   indicating that the extrinsic variation started affecting observers' lightness judgement. We estimated that
20   the variance in the lightness representation due to variation in reflectance spectra or light source intensity
21   were within a factor of two as compared to the variation in observers' intrinsic representation of object
22   lightness. Moreover, for simultaneous variation of both these spectral properties, the increase in threshold
23   square compared to no variation condition was a linear sum of the corresponding increase in threshold
24   squares for the individual properties, indicating that the noise from these independent sources combine
25   linearly.

26

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

28

INTRODUCTION

Our visual system provides perceptual representation of distal properties of objects based on the proximal stimuli captured by the eyes. While object properties are intrinsic to the object (its color, shape, etc.), the proximal stimuli also depend on the properties of the scene in which the object lies (such as background objects in the scene, illumination, etc.) as well as the position and pose of the observer. The task of the visual system is to provide stable correlates of object intrinsic properties under variability of the proximal signal due to changes in object extrinsic scene properties. This work quantifies the extent to which the visual system provides such stability for the representation of the reflectance of an object under variation in spectral properties of the scene, specifically, variation in the spectra of the background objects and the light sources in the scene.

The perceptual correlate of the reflectance of an object is its perceived color. For achromatic objects, the analogous quantity is object lightness. The human visual system is known to provide a relatively stable representation of the color/lightness of an object despite variability in the proximal signal due to changes in the surface reflectance of objects in the scene, the light source, geometry, and other properties of the scene (Foster, Color constancy., 2011; Brainard & Radonjic, Color constancy., 2004). The degree to which such stability can be achieved is termed as color/lightness constancy (Adelson, 2000; Gilchrist, 2006). Human color/lightness constancy has been measured using appearance-based approaches and discrimination-based approaches (Olkkonen M. &., 2016). Appearance based approaches involve tasks in which the observer makes judgement about the appearance of stimuli. This approach includes methods such as color matching, color naming, scaling, and nulling (Foster D. H., 2003). In color matching, observers adjust a test stimulus to match a standard stimulus. Color matching experiments show varying degrees of constancy under changes of illumination (Arend & Robert, 1987; Arend & Spehar, 1993), reflectance (Arend & Spehar, 1993), illumination gradients (Arend & Goldstein, 1990; Brainard, Brunt, & Speigle, 1997), and illumination and simulated reflectance (Rutherford & Brainard, 2002). Color naming is a more direct and arguably natural method to measure color constancy where observers are asked to categorize stimuli based on their hue, saturation, and lightness (Troost & De Weert, 1991). This method has been used with real (Uchikawa, Uchikawa, & Boynton, 1989; Olkkonen, Witzel, Hansen, & Gegenfurtner, 2010) and simulated stimuli (Maria, Hansen, & Gegenfurtner, 2009) to measure constancy, and has the limitation that there are vast number of possible discernible colors (Linhares, Pinto, & Nascimento, 2008), but there is a limit of the gamut that can be displayed. Typically, observers are asked to name from a small set of colors (Speigle & Brainard, 1996; Smithson & Zaidi, 2004; Hansen, Walter, & Gegenfurtner, 2007) which might provide an overestimate of the measured constancy. In color scaling methods, observers view a stimulus and provide a rating on a scale for a set of colors, thus allowing for a finer level of comparison for measuring constancy (Luo, et al., 1991; Schultz, Doerschner, & Maloney, 2006). Scaling methods can also be used to measure changes in stimuli, where observers provide a rating of the change between stimuli (Ennis & Doerschner, 2019). Nulling or achromatic adjustment methods involve changing a test stimulus such that it appears achromatic (Arend L. E., 1993; Brainard D. H., Color constancy in the nearly natural image. 2. Achromatic loci., 1998; Delahunt & Brainard, 2004). This method has the limitation that it provides data only for achromatic/gray stimuli and additional assumptions about the observers' criterion needs to be made for color appearances (Speigle & Brainard, 1996).

Discrimination-based approaches provide an objective method to measure color constancy (Bramwell & Hurlbert, 1996; Reeves, Amano, & Foster, 2008). In these experiments, observers discriminate stimuli to be the same or different from each other. The stimuli are varied in some relevant parameter space to measure the threshold for discriminating changes in the parameter (Craven & Foster, 1992; Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert, 2014; Aston, Radonjic, Brainard, & Hurlbert, 2019). Recently, Singh et. al (Singh, Burge, & Brainard, 2022) developed an equivalent noise paradigm that relates thresholds of discrimination to the variability in observers' intrinsic representation of object properties (e.g., its lightness) and the variability due to object extrinsic properties of the scene. They

78 measured human lightness discrimination thresholds as a function of the amount of variability in the
79 spectra of background objects in a scene. They related the discrimination thresholds to the variance in
80 observers' internal perceptual representation of lightness and the variance in the spectrally induced
81 extrinsic variability. A comparison of the strength of intrinsic and extrinsic variability provided a measure
82 of the degree of constancy in the object intrinsic property due to the variability in object extrinsic
83 property.

84 This equivalent noise paradigm can also be used to compare the effect of different sources of variabilities.
85 The variance of multiple extrinsic properties can be characterized relative to the variance of the intrinsic
86 variability. These in turn can be compared to each other to measure their relative effect. Another strength
87 of the equivalent noise paradigm is that it can be used to characterize how the effect of multiple sources
88 of variability combine when presented simultaneously.

89 In this work, we use this paradigm to compare the variation in two spectral properties of the scene to
90 human observers' representation of lightness. The spectral variations we study are: the surface reflectance
91 of the background objects in the scene and the intensity of the light sources in the scene. We measure
92 human observers' threshold of discriminating two images based on the lightness of an achromatic target
93 object in the images. We measure how these discrimination thresholds change as we increase the
94 variability in the reflectance spectra of the background objects and the intensity of the light sources. We
95 measure discrimination thresholds for individual and simultaneous variation of these properties. We use
96 the equivalent noise paradigm to relate the thresholds to the variance of observers' intrinsic noise and the
97 extrinsic variability. These variances allow one to compare the relative effect of these spectral variations.
98 A comparison of variance of individual and simultaneous variation condition can provide information
99 about the combination rules of multiple sources of variation.

100 We show that as the variability in the extrinsic sources increases, initially for small amount of variation,
101 the thresholds remain constant. In this region, the thresholds are determined primarily by the variation in
102 observers' internal representation of lightness. As the variability increases further, the discrimination
103 thresholds increase. The increase in thresholds can be accounted for by a model based on signal detection
104 theory. This model shows that the effect of extrinsic variation is within a factor of two compared to the
105 variability in the intrinsic representation. This confirms that the visual system provides a large degree of
106 lightness constancy under object extrinsic scene variability. By comparing the increase in thresholds
107 under individual and simultaneous variation from the thresholds under no extrinsic variation, we also
108 show that the effect of individual sources combines linearly under simultaneous variation.

109 The paper is organized as follows. Section 2, Experimental Methods, provides the details of the
110 experimental methods, stimuli used, and model fitting. Section 3, Results, provides the results of three
111 experiments: variation in background reflectance spectra, variation in light source intensity, and
112 simultaneous variation in these two properties. Section 4, Discussion, provides a summary of the results
113 and conclusive remarks.

2 EXPERIMENTAL METHODS

Overview

116 We followed the methodology published previously in (Singh, Burge, & Brainard, 2022). In this previous
117 work, human lightness discrimination thresholds were measured under variability of the reflectance
118 spectra of background objects in the scene. The work presented here follows the same experimental
119 methods, except that the stimuli used in the experiment are different. This section provides an overview of
120 the methods, focusing on the differences from the previous work. We refer the reader to the previous
121 work for details.

Similar to the previous work, we used a two-alternate forced-choice (2AFC) procedure to measure thresholds (Figure 1). On each trial, observers viewed two computer graphics rendered images of 3D scenes on a color calibrated monitor. Each trial contained a standard image and a comparison image. The images were viewed in sequence for 250ms with a 250ms inter-stimulus interval. Each image contained an achromatic spherical target object. The observers reported the image in which the target object was lighter. Across trials, we varied the luminous reflectance factor (LRF) of the target object (American Society for Testing and Materials, 2017) in the comparison image. The LRF is the ratio of the luminance of a surface under a reference illuminant (here, the CIE D65 reference illuminant) to the luminance of the reference illuminant itself. The order of the standard and the comparison image was chosen in a pseudorandom order. We recorded the proportion of times observers chose the comparison image as having the lighter target object at 11 values of the target object LRF. Figure 2 shows a psychometric function from a typical human observer. The proportion-comparison-chosen data were fit with a cumulative normal. Threshold was defined as the difference between the LRF of the target object at proportion comparison chosen 0.76 and 0.50 (i.e., d-prime = 1.0 in a two-interval task), as determined from the cumulative normal fit.

We measured the effect of variation in two types of object-extrinsic scene properties on human lightness discrimination thresholds: variation in the reflectance spectra of the background objects in the scene and variation in the intensity of the light sources in the scene. We performed three experiments:
(1) Background reflectance spectra variation (preregistered Experiment 6): In this experiment, we measured human lightness discrimination thresholds as a function of the amount of variation in the background objects while the spectra of the light sources were kept fixed.
(2) Light source intensity variation (preregistered Experiment 7): In this experiment, we measured lightness discrimination thresholds as a function of the amount of variation in the intensity of the light sources while the background was fixed.
(3) Simultaneous variation (preregistered Experiment 8): In this experiment, we measured lightness discrimination thresholds as both the background object reflectance spectra and the light source intensity varied simultaneously.

The reflectance spectra of the background objects were sampled from a multivariate normal distribution. The amount of variation in the spectra was controlled by multiplying the covariance matrix of the multivariate normal distribution by a scalar. By varying the covariance scalar from 0 (no variation) to 1 (natural scene variation), we studied how background reflectance affected lightness discrimination thresholds. We measured discrimination thresholds for both chromatic and achromatic variations. In chromatic variation, the reflectance spectra could take any shape and the objects varied in their luminance and chromaticity. In achromatic variation, the reflectance spectra were spectrally flat, and the objects were gray.

The shape of the spectral power distribution function of the light source was chosen as CIE D65 reference illuminant. The intensity was varied by multiplying the spectral power distribution function by a scalar sampled from a log uniform distribution. The amount of variation was controlled by changing the range of the log uniform distribution.

The subsections below provide additional methodological detail.

Preregistration

163 The experimental design and the method for extracting thresholds from the data were preregistered before
164 the start of the experiments. The preregistration documents are publicly available at: https://osf.io/7tgy8/.[1]

165 We preregistered three experiments. These were preregistered as Experiment 6 (referred here as
166 Background reflectance spectra variation), Experiment 7 (referred here as Light source intensity
167 variation), and Experiment 8 (referred here as Simultaneous variation). Experiment 6 was a replication of
168 previous work (preregistered as Experiment 3; Singh, Burge, & Brainard, 2022) with three additional
169 conditions in which the background objects were achromatic and varied only in their lightness. The
170 experimental methods of the three experiments were the same.

171 We followed the procedure described in the preregistration document to extract thresholds from the data.
172 The preregistration document also indicated that the primary data feature of interest was the dependence
173 of threshold on the amount of variation in the background reflectance and the intensity of the light source.
174 We predicted that the thresholds would increase with increase in the amount of variation. For background
175 variation, we predicted that the thresholds of achromatic variation would be lower than chromatic
176 variation. We also predicted that increase in thresholds could be captured by an equivalent noise model
177 (Singh, Burge, & Brainard, 2022). Additionally, we predicted that the threshold for simultaneous
178 variation would be higher than the thresholds for individual variations.

179 **Reflectance and Illumination Spectra**

180 The reflectance spectra of background objects in the scene were generated using a model of naturally
181 occurring surface reflectance spectra, as described in the previous work (Singh, Cottaris, Heasly,
182 Brainard, & Burge, 2018; Singh, Burge, & Brainard, 2022). Briefly, we combine two datasets of surface
183 reflectance functions containing 632 surface reflectance measurements (Vrhel, Gershon, & Iwan, 1994;
184 Kelly, Gibson, & Nickerson, 1943). We then use principal component analysis (PCA) to obtain the
185 projection of the mean centered dataset along the eigenvectors associated with the six largest eigenvalues.
186 These eigenvalues captured more than 99.5% of the variance (Singh, Cottaris, Heasly, Brainard, & Burge,
187 2018). We approximate the empirical distribution of the projection weights with a multivariate normal
188 distribution. We generate pseudorandom samples from this multivariate normal distribution to get the
189 projection weights of random samples of reflectance spectra. Reflectance spectra were constructed by
190 using these projection weights along with the eigenvectors and adding the mean of the surface reflectance
191 dataset. A physical realizability condition was imposed on these spectra by ensuring that the reflectance at
192 each wavelength was between 0 and 1. If a reflectance spectrum did not meet this criterion, it was
193 discarded.

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

197 To control the amount of variation in the reflectance spectra, the covariance matrix of the multivariate
198 normal distribution was multiplied by a covariance scalar (σ^2). A covariance scalar of 0 corresponds to
199 no variation in background object reflectance spectra. A covariance scalar of 1 corresponds to the full
200 reflectance variation of the model of natural reflectance.

---

[1] The preregistration documents relevant to this work are those for Experiments 6, 7 and 8. The site also contains preregistrations for previously reported (Experiment 1, 2 and 3; Singh, Burge, & Brainard, 2022) and unreported (Experiment 4 and 5) work.

201 The power spectrum of the light source was chosen as CIE D65 reference illuminant. We normalized the
202 D65 spectrum by its mean power to obtain its relative spectral shape. The variation in the light source
203 intensity was introduced by multiplying the normalized D65 spectrum by a random sample generated
204 from a log-uniform distribution in the range $[1-\delta, 1+\delta]$, where δ determines the range of the
205 distribution. We chose log-uniform distribution for the multiplication parameter because the spectral
206 power distribution function of natural daylight spectra varies over three orders of magnitude and their
207 mean over wavelength can be roughly approximated by a log-uniform distribution (Singh, Cottaris,
208 Heasly, Brainard, & Burge, 2018). All light sources in a scene were assigned the same power spectrum.

209 The values of the two parameters σ^2 and δ for the three experiments were as follows:

210 <u>Background object reflectance variation (Experiment 6):</u> In this experiment, we generated images for nine
211 conditions. Six of these conditions were for chromatic variation at six logarithmically spaced values of the
212 covariance scalar (σ^2): [0, 0.01, 0.03, 0.1, 0.3, 1.0]. Three conditions were for achromatic variation at
213 covariance scalar (σ^2): 0.03, 0.3 and 1.0. The power spectrum of the light source was the same for all
214 images. The power spectrum multiplication scalar was assigned an arbitrary value of 5. Figure 3 shows
215 five typical images for the nine conditions.

216 <u>Light source intensity variation (Experiment 7):</u> In this experiment, we generated images for seven
217 linearly spaced values of the range parameter (δ): [0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30]. The
218 reflectance spectra of all background objects were the same and were equal to the mean spectrum of the
219 reflectance database. This corresponds to covariance scalar of 0. Figure 4 shows five typical images for
220 the seven conditions.

221 <u>Simultaneous variation (Experiment 8):</u> In this experiment we studied six conditions. These were: no
222 variation ($\sigma^2 = 0$, $\delta = 0$), chromatic background variation (covariance scalar = 1, $\delta = 0$), achromatic
223 background variation ($\sigma^2 = 1$, $\delta = 0$), light source intensity variation ($\sigma^2 = 0$, $\delta = 0.3$), simultaneous
224 variation chromatic background ($\sigma^2 = 1$, $\delta = 0.3$) and simultaneous variation achromatic background (σ^2
225 $= 1$, $\delta = 0.3$). Figure 5 shows five typical images for these six conditions.

226 **Stimulus Design**

227 Images were generated using the software referred to as Virtual World Color Constancy (VWCC)
228 (github.com/BrainardLab/VirtualWorldColorConstancy) as described in (Singh, Burge, & Brainard,
229 2022). To render an image, we first create a 3D model as a base scene. Next, as per the experimental
230 condition, we assign reflectance spectra and spectral power distribution function to the objects and light
231 sources in the scene. All light sources in a scene are assigned the same spectral power distribution
232 function. Then we render a 2D multispectral image of the scene using Mitsuba, a physically-realistic
233 open-source rendering system (mitsuba-renderer.org; Jakob, 2010). The images were rendered at 31
234 wavelengths equally spaced between 400nm and 700nm. The images were rendered with the camera field
235 of view of 17° with an image resolution of 320-pixel by 240-pixels with the target object at the center. A
236 201-pixel by 201-pixel area, centered around the spherical target object, was cropped for display on the
237 monitor. The cropped images were converted to LMS images using the Stockman-Sharpe 2° cone
238 fundamentals (T_cones_ss2 in the Psychophysics Toolbox). Then the monitor calibration data and
239 standard methods (Brainard D. H., 1989; Brainard, Pelli, & Robson, 2002) were used to convert the LMS
240 images to gamma corrected RGB images. A common scaling was applied to all images before rendering
241 to ensure that they were within the monitor's gamut, so that the maximum linear channel RGB channel
242 input was 0.9. The gamma corrected RGB images were presented on the monitor during the experiment.

243 For each condition described above, we generated 1100 images, 100 images each at each of the 11
244 linearly spaced values of the target object LRF in the range [0.35, 0.45]. The standard image target object
245 LRF was 0.4. The comparison image target object LRF varied in the range [0.35, 0.45]. We generated 100
246 images at each comparison level to avoid excessive replication of images in the experiment. For the no
247 variation ($\sigma^2 = 0$, $\delta = 0$) condition, we generated one image at each target object LRF level, as the
248 background reflectance and the light source intensity remained fixed in this case. The images were
249 generated without secondary reflections. The geometry of the scene was also fixed.

250 When displayed on the experimental monitor, the average luminance of the standard image for $\sigma^2 = 0.00$
251 and $\delta = 0.00$ was 87.1 cd/m^2. The average luminance of the target object for the 11 LRF levels were
252 [120.9, 122.3, 123.8, 125.2, 126.5, 127.9, 129.2, 130.5, 131.9, 133.1, 134.4] cd/m^2.

253 When displayed on the experimental monitor, the average luminance of the standard image for $\sigma^2 = 1.0$
254 and $\delta = 0.30$ was 87.8 cd/m^2. The average luminance of the target object for the 11 LRF levels were
255 [117.7, 119.4, 119.4, 122.3, 123.7, 123.8, 127.8, 126.9, 127.7, 129.1, 129.0] cd/m^2.

256 **Experimental Structure:**

257 We define a trial as the presentation of a standard and a comparison image and the recording of the
258 response. We define an interval as the presentation of one of the images, standard or comparison, in a
259 trial. A block consists of recording 330 trials for one condition, 30 trials each at 11 comparison image
260 target LRF levels. A permutation consists of recording one block of data for each condition in an
261 experiment. We recorded three permutations for each observer in each experiment. Each permutation had
262 a random order of the conditions.

263 The order of the blocks in a permutation, the LRF levels of the comparison image in trials of a block, and
264 the order of standard and comparison images in a trial was generated pseudorandomly and stored at the
265 beginning of the experiment for each observer. Before starting a new permutation for an observer, the
266 data for all conditions in a permutation was collected.

267 A session consisted of recording three blocks on a single day. An observer performed no more than one
268 session on a day. Each block in a session was divided into three sub-blocks of 110 trials. Between these
269 sub-blocks, the observers took a break of minimum one minute. The observers also took a small break
270 (two to five minutes) between blocks. The observers could terminate the experiment at any time during
271 the block. If the observer terminated a block, the data was not recorded. No observer terminated a block
272 of the experiment.

273 Each observer first performed a practice session where three blocks of data was recorded for the no
274 variation ($\sigma^2 = 0$, $\delta = 0$) condition. The observers were excluded from the experiment if their mean
275 threshold for the last two blocks was higher than 0.03. If the observer passed this criterion, then the rest of
276 the data was collected over several days.

277 At the beginning of the practice session for each observer, the experimenter explained the experimental
278 procedure and obtained consent for the experiment. Then the observer was tested for normal visual acuity
279 and color vision. After this, the observer went to the experimental room where they were familiarized
280 with the experimental set-up by performing a familiarization block of 40 trials. Then the observers were
281 dark adapted by sitting in the dark for about 5 minutes. Then the data for the three blocks of the practice
282 session was recorded. At the end of the practice session, the observers were informed if they could
283 continue the experiment.

284 If the observer was continued, their data was collected over several sessions. The data for all six observers
285 of an experiment was collected over several weeks. The data of all six observers for preregistered
286 Experiment 6 was collected before starting preregistered Experiment 7. The data of all six observers for
287 preregistered Experiment 7 was collected before starting preregistered Experiment 8.

Observer Recruitment and Exclusion

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

294 Observers who passed the vision screening then participated in a practice session. This session also served
295 to screen for the observers' ability to reliably perform the psychophysical task. In the practice session, the
296 observers' performed three blocks of the experiment for the no variation condition ($\sigma^2 = 0.00$, $\delta = 0.00$)
297 and the threshold was calculated for these three blocks. The observer was excluded from further
298 participation if their mean threshold for the last two blocks in the practice session exceeded 0.030 (log T^2,
299 -3.2). This exclusion criterion was specified in our preregistered protocol (See Methods: Preregistration).

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

301 For each observer, the practice session was performed at the beginning of each of the three experiments
302 (Experiment 6, 7, 8), irrespective of whether the observer had participated in an earlier experiment.

Observer Information

304 Background reflectance variation (preregistered Experiment 6): A total of 25 observers participated in the
305 practice sessions for background variation experiment (10 Female, 15 Male; age 19-34; mean age 22.9).
306 To de-identify observer information in the data, observers were given pseudo-names chosen by the
307 experimenter. Six of these observers (pseudo-names: *0003, bagel, committee, content, observer*, and
308 *revival*) met the performance criterion set for screening (2 Female, 4 Male; age 19-28; mean age 23.33).
309 All observers who advanced to the practice session had normal or corrected-to-normal vision (20/40 or
310 better in both eyes, assessed using Snellen chart) and normal color vision (0 Ishihara plates read
311 incorrectly). The visual acuities of the observers in the main experiment were: *0003*, L = 20/30, R =
312 20/20; *bagel*, L = 20/20, R = 20/20; *committee*, L = 20/25, R = 20/25; *content*, L = 20/20, R = 20/20;
313 *observer*, L = 20/25, R = 20/25; *revival*, L = 20/20, R = 20/20. *Committee, content,* and *observer* wore
314 personal corrective eyewear both during vision testing and during the experiments. Observers *0003,*
315 *bagel*, and *revival* did not require or use corrective eyewear.

316 Light source intensity variation (preregistered Experiment 7): A total of 15 observers participated in the
317 practice sessions for light source intensity variation experiment (9 Female, 6 Male; age 19-33; mean age
318 25). Six of these observers (pseudo-names: *0003*, *bagel*, *content*, *oven*, *primary*, and *revival*) met the
319 performance criterion set for screening (3 Female, 3 Male; age 19-28; mean age 23.83). All observers
320 who advanced to the practice session had normal or corrected-to-normal vision (20/40 or better in both
321 eyes, assessed using Snellen chart) and normal color vision (0 Ishihara plates read incorrectly). The visual
322 acuities of the observers in the main experiment were: *0003*, L = 20/30, R = 20/30; *bagel*, L = 20/20, R =
323 20/20; *content*, L = 20/20, R = 20/20; *oven*, L = 20/20, R = 20/20; *primary*, L = 20/20, R = 20/20; *revival*,
324 L = 20/20, R = 20/20. Observer *content* and *primary* wore personal corrective eyewear both during vision
325 testing and during the experiments. Observers *0003, bagel*, *oven,* and *revival* did not require or use
326 corrective eyewear. Observer *oven* reported some difficulties during a few sessions of the experiment and
327 their thresholds for two conditions did not fit the expected pattern. We removed their data from the
328 analysis presented in this work. Their data and thresholds are provided in supplementary materials.

329   Simultaneous variation (preregistered Experiment 8): A total of 20 observers participated in the practice
330   sessions for simultaneous variation experiment (9 Female, 11 Male; age 19-28; mean age 20.8). Six of
331   these observers (pseudo-names: *0003, bagel, content, oven, manos,* and *revival*) were retained for the
332   experiment (2 Female, 4 Male; age 19-28; mean age 23.33). Four observers (*0003, bagel, content,* and
333   *oven*) met the screening criteria specified in the preregistration. Due to lack of observers who met the
334   preregistration criteria, two observers (*manos,* and *revival*), whose thresholds were close to the
335   preregistration criteria, were also retained for the experiment. Observer *revival* had participated in
336   previous two experiments and had met the criteria both times. Observer *manos* showed improvement in
337   thresholds with each block, with the threshold for the final block below 0.03. This was a deviation from
338   the preregistration. All observers who advanced to the practice session had normal or corrected-to-normal
339   vision (20/40 or better in both eyes, assessed using Snellen chart) and normal color vision (0 Ishihara
340   plates read incorrectly). The visual acuities of the observers in the main experiment were: *0003*, L =
341   20/30, R = 20/30; *bagel*, L = 20/20, R = 20/20; *content*, L = 20/20, R = 20/20; *oven*, L = 20/20, R =
342   20/20; *manos*, L = 20/25, R = 20/25; *revival*, L = 20/20, R = 20/20. Observer *content* wore personal
343   corrective eyewear both during vision testing and during the experiments. Observers *0003, bagel*, *manos*,
344   *oven*, and *revival* did not require or use corrective eyewear.

Apparatus

346   The stimuli were presented on a calibrated LCD color monitor (27-in. NEC MultiSync EA271U; NEC
347   Display Solutions) in an otherwise dark room. The monitor was driven at a pixel resolution of 1920 x
348   1080, a refresh rate of 60Hz, and with 8-bit resolution for each RGB channel. The host computer was an
349   Apple Macintosh with an Intel Core i7 processor. The experimental programs were written in MATLAB
350   (MathWorks; Natick, MA) and relied on routines from the Psychophysics Toolbox
351   (http://psychtoolbox.org) and mgl (http://justingardner.net/doku.php/mgl/overview). Responses were
352   collected using a Logitech F310 gamepad controller.

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

Monitor Calibration

357   The monitor was calibrated using a spectroradiometer (PhotoResearch PR655) as described in (Singh,
358   Burge, & Brainard, 2022). The monitor was calibrated before starting each experiment. Once calibrated,
359   the same settings were used till data for all observers for that experiment was collected. The monitor was
360   then recalibrated for the next experiment. Data was collected in the sequence Experiment 6, Experiment
361   7, and Experiment 8.

Ethics Statement

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

Code and Data Availability

367   For each experiment and observer, the proportion comparison chosen data for the 3 experimental blocks
368   of each condition as well as the thresholds are provided as supplementary information (SI). The SI also
369   provides the MATLAB scripts to generate Figures 2, 6 – 12, supplementary figures S1-S5, and the scripts

370 to obtain thresholds of the linear receptive field formulation of the model. The SI is available at:
371 https://github.com/vijaysoophie/SimultaneousVariationPaper.

Linear Receptive Field Model

373 The thresholds of preregistered Experiment 6 were fit to the linear receptive field model developed in
374 (Singh, Burge, & Brainard, 2022). This model consisted of a simple center surround receptive field (R).
375 The receptive field was square in shape to match the images in the psychophysics experiment. Its center
376 was a circle of radius equal to the size and the location of the target object. The central region had a
377 spatially uniform positive sensitivity of 1. The surround had a spatially uniform negative sensitivity of v_s.
378 The receptive field response was computed as the dot product of the receptive field with the standard and
379 the comparison images. A mean zero Gaussian noise was added to the response. The image with the
380 higher noise added receptive field response was chosen to be lighter. The variance of the Gaussian noise
381 (σ_{ri}^2) and the value of the receptive field surround sensitivity (v_s) were the two parameters of the model.
382 These parameters provided an estimate of the internal noise (σ_{ri}^2) and the variance of the extrinsic
383 properties (σ_{e0}^2). The model related the thresholds (T) in the experiments to the variance in the intrinsic
384 noise (σ_{ri}^2) of the observer and the extrinsic variance (σ_{e0}^2) through the relation:

$$385 \qquad T = \sqrt{\sigma_{ri}^2 + \sigma^2 \ \sigma_{e0}^2} \qquad\qquad (1)$$

386 where σ^2 is the covariance scalar (see (Singh, Burge, & Brainard, 2022) for details). The variance of the
387 extrinsic properties (σ_{e0}^2) resulting from the variation in the image can be computed as $R^T \Sigma_{e0} R$, where
388 Σ_{e0} is the covariance matrix of the variation in the images.

389 <u>Background reflectance variation:</u> To estimate the variance of the intrinsic noise of the observer and
390 extrinsic variation in the images, we chose the value of the Gaussian noise variance ($\sigma_{ri,B}^2$) and the
391 surround sensitivity $v_{s,B}$ to minimize mean squared difference between the model and experimental
392 thresholds measured at the six values of the covariance scalar. $\sigma_{ri,B}^2$ provided the estimate of the intrinsic
393 noise. The extrinsic noise ($\sigma_{e0,B}^2$) was estimated by using the best fit surround sensitivity ($v_{s,B}$) of the
394 receptive field (R) and the sample covariance matrix of the images (Σ_{e0}) at $\sigma^2=1$ in the relation $R^T \Sigma_{e0} R$.

395 <u>Light source intensity variation:</u> We fit a functional form similar to Eq. (1) to the thresholds of light
396 source intensity variation experiment (Experiment 7) where we replaced covariance scalar σ^2 by the
397 range parameter δ. We chose the value of the Gaussian noise variance ($\sigma_{ri,L}^2$) and receptive field surround
398 sensitivity (v_s) to minimize the mean square difference between the observer and model thresholds
399 measured at the seven values of the range parameter. $\sigma_{ri,L}^2$ provided the estimate of the observer's intrinsic
400 noise. To estimate the extrinsic noise, we used the best fit surround sensitivity ($v_{s,L}$) and the sample
401 covariance matrices to calculate the quantity $R^T \Sigma_{e0} R$ as a function of the range parameter δ. We fit the
402 resulting values with an exponential function (see Figure S1). The extrinsic noise ($\sigma_{e0,L}^2$) was estimated as
403 the value of the exponential fit at $\delta = 1$.

404 **3 RESULTS**

Human Lightness Discrimination Thresholds Increase with Background Reflectance Variation

406 We measured lightness discrimination thresholds of human observers for two types of variation in the
407 reflectance spectra of background objects in the scene: chromatic variation and achromatic variation. In

408 chromatic variation, the reflectance spectra could take any shape and thus the background objects varied
409 in their chromaticity and luminance. In achromatic variation, each spectrum had the same reflectance at
410 all wavelengths, and thus the spectra varied only in their overall luminance and the objects were gray. The
411 amount of variation depended on the covariance matrix of the multivariate normal distribution from
412 which the spectra were sampled. The variance was controlled by multiplying the covariance matrix by a
413 covariance scalar (σ^2). We measured discrimination thresholds of six human observers at six values of
414 the covariance scalar for chromatic variation and three values of covariance scalar for achromatic
415 variation. The threshold was measured three times (three separate blocks) for each observer and for each
416 of the nine conditions. The psychometric functions for these nine conditions are shown for one observer
417 in Figure 6 and for all observers in Figure S2. Inspection of the psychometric functions show that their
418 slopes steadily decrease with increasing covariance scalar, corresponding to an increase in thresholds. The
419 thresholds for chromatic and achromatic variation are comparable.

420 Figures 7 shows explicitly how the discrimination thresholds change with the amount of variability in the
421 reflectance of the background objects. Here, we plot the mean log threshold squared (averaged across
422 observers, N = 6) against the log of the covariance scalar. Table S1 provides the thresholds and SEMs
423 from Figure 7 in tabular form. For low values of the covariance scalar, the thresholds are nearly constant.
424 As the covariance scalar increases, log threshold squared increases. The thresholds are comparable for
425 chromatic and achromatic variation. The p-values of the hypothesis that the mean thresholds for
426 chromatic and achromatic variations are equal are 0.72, 0.57, and 0.16 for covariance scalar 0.03, 0.30,
427 and 1.00 respectively, indicating that the differences in the mean thresholds are not statistically
428 significant.

429 We fit the thresholds to the linear receptive field (LINRF) model (Eq. 1) developed in (Singh, Burge, &
430 Brainard, 2022). The LINRF model provides the estimate of the variance of the internal noise of the
431 observer as $\sigma^2_{\text{ri,B}} = 0.026$ and the variance of the extrinsic variability due to the reflectance of
432 background objects as $\sigma^2_{\text{e0,B}} = 0.039$. The equivalent noise level, the ratio of the external variance to
433 intrinsic noise, is ~ 1.5, indicating that the variability in the representation of object lightness induced by
434 the natural variability in the reflectance of background objects is close to the internal variability of that
435 representation. If the ratio was equal to 1, then we would have concluded that the visual system has
436 discounted the external variability. But the ratio not significantly large compared to 1, this indicates that
437 the visual system provides a significant level of lightness constancy.

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

439 We measured lightness discrimination thresholds of human observers as we varied the intensity of light
440 sources in the scene. The shape of the spectrum of the light sources was fixed to be standard daylight
441 spectrum D65. We normalized the spectrum by its mean over wavelengths. The intensity was varied by
442 multiplying the normalized spectrum by a scalar sampled from a log-uniform distribution in the range [1-
443 δ, 1+ δ]. The reflectance spectra of the background objects were fixed. We measured lightness
444 discrimination thresholds for seven values of the range parameter δ for five human observers. The
445 psychometric function of one of the observers for these seven conditions are shown in Figure 8. Figure S3
446 shows the psychometric functions of all observers. Figure 9 shows how the thresholds change as the
447 amount of variation in the light source intensity increases. The data is averaged over five observers (also
448 see Figure S4). Similar to the trend for reflectance spectra variation, lightness discrimination thresholds
449 remain constant for small values of the range parameter and then log threshold squared increases with
450 increase in the range parameter. A fit of the mean squared threshold with the linear receptive field model
451 gives the value of internal noise as $\sigma^2_{\text{ri,L}} = 0.028$. This compares well with the internal noise obtained
452 from the background reflectance spectra variation experiment ($\sigma^2_{\text{ri,B}} = 0.026$). The variance of the
453 extrinsic variability is $\sigma^2_{\text{e0,L}} = 0.052$. The equivalent noise level is ~ 1.8, indicating that the variation in

454 the lightness representation induced by the variation in light source intensity is close to the internal
455 variation of that representation.

Thresholds for Simultaneous Variation are Higher Than Individual Variations

457 We measured lightness discrimination thresholds for simultaneous variation in the reflectance spectra of
458 background objects and the intensity of the light sources in the scene. In this experiment, we studied six
459 conditions: no variation, variation in the reflectance spectra of background objects with fixed spectrum of
460 the light sources for achromatic and chromatic backgrounds, variation in intensity of light source with
461 fixed background, and simultaneous variation in the intensity of light source and reflectance spectra of
462 background object for chromatic and achromatic backgrounds. We measured lightness discrimination
463 thresholds of six human observers for these six conditions. The psychometric function of one of the
464 observers is shown in Figure 10. Figure S5 shows the psychometric functions of all observers. Figure 11
465 shows the mean log squared threshold of all six observers for these six conditions. The threshold for
466 simultaneous variation of light source intensity and reflectance spectra of background objects is higher
467 than the condition with individual variations in these properties. As observed earlier, the threshold for
468 achromatic and chromatic conditions are comparable. The p-value of the hypothesis that the mean
469 thresholds for chromatic and achromatic variations are equal is 0.19 for background variation condition
470 and 0.44 for simultaneous variation condition, indicating that the differences in the mean thresholds are
471 not statistically significant.

472 Figure 11 also shows the squared thresholds of the linear receptive model for the six conditions. We used
473 the intrinsic noise and the surround sensitivity (v_s) parameters of the background reflectance variation
474 experiment to estimate the threshold of the linear receptive field model for the no-variation condition,
475 background spectra variation conditions, and the simultaneous variation conditions (Experiment 6, Figure
476 7). For the light intensity variation condition, we used the parameters of the light source intensity
477 variation experiment (Experiment 7, Figure 9).

478 We can use the linear receptive field model to compare the extrinsic variance of the simultaneous
479 variation condition to the variance of the individual variations. According to the linear receptive model,
480 the square of the threshold is proportional to the sum of the variance of observers' intrinsic noise and the
481 extrinsic variation in the stimuli (Eq. 1). The squared threshold at the no variation condition is equal to the
482 variance of the observers' intrinsic noise. In case of extrinsic variation, the increase in threshold square
483 compared to the no variation condition equals the variance of the extrinsic variation. When there are more
484 than one independent sources of extrinsic variation, the total variance of the simultaneous variation
485 should be the sum of the variance of the individual variations. This predicts that the increase in threshold
486 square for simultaneous variation condition should be equal to the sum of the corresponding increase for
487 the individual variation conditions.

488 Figure 12 shows the increase in mean squared threshold above the no variation condition. We compare
489 the mean square thresholds of the simultaneous variation condition with the sum of the mean square
490 thresholds of the individual conditions for chromatic and achromatic conditions. The increase in threshold
491 of the simultaneous variation condition is comparable to the sum of the increase in threshold for the
492 individual variations. The p-value of the hypothesis that the mean increase in thresholds for simultaneous
493 variation is equal to sum of the mean increase in the thresholds of light intensity variation and background
494 object reflectance variation are 0.86 and 0.80 for chromatic and achromatic conditions respectively. The
495 variance of the extrinsic noise calculated for the background variation condition ($\sigma^2 = 1, \delta = 0$) is
496 0.0015 and the light intensity variation condition ($\sigma^2 = 0, \delta = 0.3$) is 0.0017. As expected, the variance
497 of the simultaneous variation condition ($\sigma^2 = 1, \delta = 0.3$), which is 0.0033, is comparable to the sum of
498 individual variances (0.0032).

4 DISCUSSION

The perceived lightness of an object depends on the scene in which it lies. Although the variability in object-extrinsic properties of the scene causes variability in the proximal signal to the visual system, the visual system provides a relatively stable representation of the object lightness. We measured human observers' threshold of discriminating two objects based on their lightness as a function of amount of variability in the spectra of background objects and light sources in a scene. For low levels of variability, the thresholds first remained constant, showing that in this regime the performance was determined by observers' intrinsic noise. As the variability increased, the effect of extrinsic variation started dominating the performance and the discrimination thresholds increased. Using a model based on signal detection theory, we related the thresholds in the low variability regime to the internal noise of the observer. The model also related the increase in threshold to amount of variability in the extrinsic property, thus providing a comparison of the variance in extrinsic property to the intrinsic noise. The effect of both types of extrinsic variation, spectra of background objects and intensity of light sources, were comparable to the effect of intrinsic noise, showing that the visual system provides a good degree of constancy to these variations. Further, for simultaneous variation of these properties the effects added linearly, resulting in the variance of the simultaneous variation to be equal to the sum of the variance of the individual variations.

Chromatic v/s Achromatic Variations: Lightness discrimination thresholds of chromatic and achromatic variation in the reflectance spectra of background objects were statistically similar. The chromatic aspect of the variation does not seem to influence lightness discrimination, indicating that lightness and chromaticity are encoded independently. This hypothesis could be tested by measuring chromaticity discrimination thresholds under chromatic and achromatic variation of background objects.

Visual system at threshold level: For the spectral variabilities studied in this work, the variances in observers' representation of lightness due to extrinsic variations are within a factor of two compared to the variances in observers' intrinsic representation of lightness. If these variances were equal, one could conclude that the visual system has fully compensated for the extrinsic variation. As the extrinsic variances are larger than the variance of the intrinsic noise, the visual system has not fully compensated the external variabilities. But since these variances are within a factor or two, it shows that the visual system provides a large degree of stability in the perceptual representation of lightness and seems to work at near threshold levels.

Rules of Combination: The increase in squared threshold of simultaneous variation of reflectance spectra of background object and intensity of light sources from no variation condition were equal to the sum of the increase in squared threshold of the individual variations. This could be accounted assuming that the sources of noise are independent, and their effect add linearly.

Table S1: Thresholds for Background Variation Experiment (Preregistered Experiment 6):
Mean threshold (averaged over blocks) $\pm$ SEM of six human observers for nine background variation conditions studied in preregistered experiment 6.

Condition	Observer					
	0003	Bagel	Committee	Content	Observer	Revival
$\sigma^2 = 0.00$	0.0221±0.0010	0.0185±0.0018	0.0344±0.0027	0.0223± 0.0012	0.0311±0.0053	0.0251±0.0023
$\sigma^2 = 0.01$	0.0215±0.0009	0.0194±0.0020	0.0386±0.0103	0.0193±0.0012	0.0263±0.0059	0.0262±0.0048
$\sigma^2 = 0.03$	0.0242±0.0019	0.0261±0.0020	0.0285±0.0029	0.0246±0.0046	0.0292±0.0007	0.0282±0.0016
$\sigma^2 = 0.03$ Achromatic	0.0255±0.0019	0.0213±0.0024	0.0343±0.0055	0.0227±0.0023	0.0267±0.0040	0.0263±0.0016
$\sigma^2 = 0.10$	0.0278±0.0015	0.0238±0.0010	0.0284±0.0017	0.0278±0.0035	0.0335±0.0024	0.0281±0.0013
$\sigma^2 = 0.30$	0.0348±0.0025	0.0277±0.0024	0.0344±0.0020	0.0286±0.0002	0.0277±0.0019	0.0301±0.0038
$\sigma^2 = 0.30$ Achromatic	0.0333±0.0032	0.0284±0.0028	0.0319±0.0047	0.0308±0.0015	0.0358±0.0030	0.0287±0.0022
$\sigma^2 = 1.00$	0.0416±0.0072	0.0316±0.0008	0.0379±0.0024	0.0323±0.0022	0.0405±0.0042	0.0360±0.0055
$\sigma^2 = 1.00$ Achromatic	0.0289±0.0017	0.0310±0.0015	0.0391±0.0029	0.0384±0.0058	0.0312±0.0015	0.0322±0.0009

Table S2. Thresholds for Lightness Intensity Variation Experiment (Preregistered Experiment 7):
Mean threshold (averaged over blocks) $\pm$ SEM of six human observers measured for seven lightness intensity conditions studied in preregistered experiment 7. The thresholds of observer Oven were not used in Figure 9.

Condition	Observer					
	0003	Bagel	Content	Oven	Primary	Revival
$\delta = 0.00$	0.0217±0.0012	0.0181±0.0001	0.0208±0.0014	0.0520±0.0114	0.0329±0.0061	0.0372±0.0008
$\delta = 0.05$	0.0228±0.0018	0.0229±0.0018	0.0207±0.0007	0.0580±0.0064	0.0346±0.0042	0.0364±0.0013
$\delta = 0.10$	0.0275±0.0024	0.0217±0.0009	0.0242±0.0040	0.0325±0.0022	0.0343±0.0013	0.0376±0.0072
$\delta = 0.15$	0.0316±0.0009	0.0238±0.0011	0.0323±0.0032	0.0333±0.0019	0.0345±0.0042	0.0326±0.0002
$\delta = 0.20$	0.0447±0.0100	0.0381±0.0046	0.0276±0.0016	0.0493±0.0120	0.0423±0.0050	0.0392±0.0034
$\delta = 0.25$	0.0433±0.0052	0.0393±0.0062	0.0308±0.0023	0.0461±0.0060	0.0532±0.0083	0.0387±0.0025
$\delta = 0.30$	0.0404±0.0018	0.0429±0.0033	0.0347±0.0014	0.0580±0.0061	0.0465±0.0047	0.0421±0.0042

541 **Table S3. Thresholds for Simultaneous Variation Experiment (Preregistered Experiment 8)**:
542 Mean threshold (averaged over blocks) $\pm$ SEM of six human observers measured for six conditions studied in
543 preregistered experiment 8.

Condition	Observer					
	0003	Bagel	Content	Oven	Manos	Revival
No Variation	0.0261$\pm$0.0022	0.0227$\pm$0.0019	0.0246$\pm$0.0004	0.0383$\pm$0.0066	0.0258$\pm$0.0036	0.0366$\pm$0.0085
Background Variation Chromatic	0.0414$\pm$0.0036	0.0340$\pm$0.0058	0.0392$\pm$0.0083	0.0498$\pm$0.0050	0.0306$\pm$0.0013	0.0383$\pm$0.0033
Background Variation Achromatic	0.0394$\pm$0.0027	0.0319$\pm$0.0015	0.0427$\pm$0.0074	0.0683$\pm$0.0048	0.0435$\pm$0.0071	0.0389$\pm$0.0010
Light Intensity Variation	0.0464$\pm$0.0027	0.0656$\pm$0.0208	0.0412$\pm$0.0021	0.0592$\pm$0.0091	0.0464$\pm$0.0046	0.0474$\pm$0.0069
Simultaneous Variation Chromatic	0.0635$\pm$0.0092	0.0536$\pm$0.0014	0.0437$\pm$0.0011	0.0639$\pm$0.0106	0.0768$\pm$0.0085	0.0528$\pm$0.0037
Simultaneous Variation Achromatic	0.0648$\pm$0.0103	0.0540$\pm$0.0017	0.0478$\pm$0.0049	0.0826$\pm$0.0166	0.0749$\pm$0.0082	0.0561$\pm$0.0028

544

Figure Captions

Figure 1: (a) Psychophysical task. On every trial of the experiment, human observers viewed two images, a standard image and a comparison image, and indicated the image in which the spherical target object at the center of the image was lighter. The images were computer graphics renderings of 3D scenes. They were displayed on a color calibrated monitor. This panel shows examples of standard and comparison images. The reflectance spectrum of the target object was spectrally flat, and the target object appeared gray. The reflectance of the target object in the standard image was held fixed and it changed for the comparison image. In this panel, the target object in the comparison image is lighter. We measured the fraction of times the observers chose the target object in the comparison image to be lighter as a function of the lightness of the target object in the comparison image. Fraction comparison chosen data was used to determine lightness discrimination threshold (Figure 2). We studied how the lightness discrimination thresholds changed as the trial-to-trial variability in the reflectance spectra of the background objects and the intensity of the light sources increased. **(b) Trial sequence:** R_{N-1} indicates the recording of the observer's response for the $(N-1)^{th}$ trial. The Nth trial begins 250ms after the completion of the $(N-1)^{th}$ trial (Inter Trial Interval, ITI = 250ms). In the Nth trial, the standard and comparison images are presented for 250ms each with a 250ms inter stimulus interval (ISI) in between the two images. The order of the standard and comparison images is chosen in pseudorandom order. The observer records their choice by pressing a button on a gamepad after both images have been presented and removed from the screen. The observers could take as long as they wish before making their choice. The recording of their choice is indicated by R_N in the panel. The next trial begins 250ms after the choice has been recorded.

Figure 2: Psychometric function: We recorded the proportion of times the observers chose the target in the comparison image to be lighter as a function of the LRF of the target object in the comparison image. We collected 30 responses each at 11 linearly spaced values of the comparison image target object LRF in the range [0.35, 0.45]. The LRF of the target object in the standard image was 0.40. The LRF of the target object in the comparison image was chosen in a pseudorandom order. The proportion comparison chosen data was fit by a cumulative normal distribution using maximum likelihood methods. The guess rate and lapse rate were constrained to be equal and 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 0.50 as obtained from the cumulative normal fit. This figure shows the data for observer 0003 in the second block of background reflectance variation experiment (preregistered Experiment 6) for the no variation ($\sigma^2 = 0.00$) condition. The discrimination threshold was 0.0208. The point of subjective equality (PSE, the LRF at which proportion comparison chosen is 0.5) was 0.409. The lapse rate for this fit was 0.00.

Figure 3: Background object reflectance variation: We studied two types of variations in the reflectance spectra of background objects in the scene: chromatic variation and achromatic variation. In chromatic variation, the reflectance spectra could take any shape, and the objects varied in their luminance and chromaticity. In achromatic variation, the reflectance spectra were spectrally flat, and the objects appeared gray and varied only in their luminance. The spectra 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 logarithmically spaced values of the covariance scalar for chromatic variation and at three values of the covariance scalar for achromatic variations. The figure shows five typical images for each of these nine conditions. For each condition we generated 1100 images, 100 images at 11 linearly spaced value of target object LRF in the range [0.35, 0.45]. The target object in each image in the figure is at LRF = 0.4.

Figure 4: Light intensity variation: The shape of the power spectrum of the light sources in the scene was chosen to be CIE reference illuminant D65. The intensity of the power spectrum was varied by multiplying the normalized D65 spectrum with a scalar sampled from a log uniform distribution in the

16

592   range [1- δ, 1+ δ]. The amount of variation was controlled by changing the value of the range parameter
593   δ. We generated images at seven linearly spaced values of the range parameter in the range [0.00, 0.30].
594   For each value of the range parameter, we generated 1100 images, 100 images at each value of the target
595   object LRF in the range [0.35, 0.45]. The figure shows five sample images at each of the seven values of
596   the range parameter. The target object in each image in the figure has the same LRF of 0.40.

597   **Figure 5: Simultaneous variation:** This figure shows five sample images for the six conditions studied
598   in preregistered experiment 8. We generated 1100 images for each of these conditions, 100 images at each
599   value of the target object LRF in the range [0.35, 0.45].

600   **Figure 6: Psychometric functions for observer 0003 for background reflectance variation**
601   **experiment:** We measured the proportion comparison chosen data for the nine conditions separately in
602   three blocks for each observer. The data for each block was fit with a cumulative normal to obtain the
603   discrimination thresholds (see Figure 2). Each panel plots the measured values and the cumulative fit to
604   the proportion comparison data for each of the three blocks, for observer 0003. The psychometric
605   functions for all six observers are shown in Figure S2. The values in the legend provide the estimate of
606   lightness discrimination threshold for each block obtained from the cumulative fit. The top row shows the
607   data for chromatic variation conditions. The last three panels in the bottom row show the data for the
608   three achromatic conditions. The first panel in the bottom row shows the data and thresholds for the
609   selection session. The selection session was a practice session in which the thresholds for the no variation
610   condition was measured three times. An observer was selected for the experiment only if the average of
611   their last two discrimination threshold measurements in the selection session was less than 0.30.

612   **Figure 7: Background variation increases lightness discrimination thresholds.** Mean (N = 6) log
613   squared threshold vs log covariance scalar from human psychophysics for chromatic (red circles) and
614   achromatic conditions (gray diamonds). The error bars represent +/- 1 SEM taken between observers. The
615   threshold of the linear receptive field (LINRF) model was estimated by simulation for the six values of
616   the covariance scalar (blue squares). The blue error bars show +/- 1 standard deviation estimated over 10
617   independent simulations. The parameters of the LINRF fit are provided in the legend. The data has been
618   jittered for ease of viewing.

619   **Figure 8: Psychometric functions for observer 0003 for light intensity variation experiment:** Same
620   as Figure 6, but for the light intensity variation experiment. The figure shows the proportion comparison
621   chosen data for the selection session and the seven condition for observer 0003. The psychometric
622   functions for all observers are shown in Figure S3.

623   **Figure 9: Light source intensity variation increases lightness discrimination threshold.** Mean (N = 5)
624   log squared threshold vs range parameter from human psychophysics for the seven light source intensity
625   variation conditions (red circles). The error bars represent +/- 1 SEM taken between observers. The
626   threshold of the linear receptive field (LINRF) model was estimated by simulation for the seven values of
627   the range parameters (blue squares). The blue error bars show +/- 1 standard deviation estimated over 10
628   independent simulations. The parameters of the LINRF fit are provided in the legend. The data has been
629   jittered for ease of viewing.

630   **Figure 10: Psychometric functions for observer 0003 for simultaneous variation experiment:** Same
631   as Figure 6 and 8, but for simultaneous variation experiment. The figure shows the proportion comparison
632   chosen data for the selection session and the six condition for observer 0003. The data for all observers
633   are shown in Figure S5.

634   **Figure 11: Discrimination thresholds for simultaneous variation of two sources are higher than**
635   **individual discrimination thresholds.** Mean (N = 6) log squared threshold for the six conditions in

17

636 simultaneous variation experiment. The error bars represent +/- 1 SEM taken between observers. The data
637 for chromatic (red circles) and achromatic (gray diamonds) conditions have been plotted next to each
638 other for visual comparison. The thresholds of the linear receptive field (LINRF) model (blue squares)
639 were estimated using the parameters of the background variation condition (Figure 7) for the None,
640 Background and Simultaneous conditions and using the parameters of the light intensity variation
641 condition for (Figure 9) for the Light condition. The blue error bars show +/- 1 standard deviation
642 estimated over 10 independent simulations.

643 **Figure 12: Extrinsic noise of independent variations add linearly for simultaneous variation:** Mean
644 squared thresholds (N=6) for the six conditions in simultaneous variation experiment (black circles). The
645 black error bars represent +/- 1 SEM taken between observers. The bars (red, gray, blue) represent the
646 increase in squared thresholds compared to the no variation condition (blue dotted line). For the
647 simultaneous variation conditions, the bars on the right (bars with one color, red or gray) represent the
648 increase in measured squared threshold for the simultaneous variation conditions and the bars on the left
649 (stacked bars of two different colors) represent the increase in the sum of the squared threshold of the
650 light intensity variation (blue bar) and the corresponding background variation conditions (red or gray).

651 **Figure S1: Estimation of extrinsic noise for light intensity variation:** Plot of the variance ($R^T \Sigma R$) as a
652 function of the range parameter δ on a linear (left panel) and logarithmic (right panel) scale. We fit the
653 function with an exponential of the form $A * \exp(B \cdot \delta)$. The variance in the extrinsic noise is estimated
654 as the value of the fit at $\delta = 1$.

655 **Figure S2: Psychometric functions for all observers for background variation experiment.** Same as
656 Figure 6, for all observers retained in background variation experiment.

657 **Figure S3: Psychometric functions for all observers for light intensity variation experiment.** Same
658 as Figure 8, for all observers retained in light intensity variation experiment.

659 **Figure S4:** Same as Figure 9, for all six observers in retained in light intensity variation experiment. The
660 parameters for the LINRF model are the same as in Figure 9.

661 **Figure S5: Psychometric functions for all observers for simultaneous variation experiment.** Same as
662 Figure 10, for all observers retained in simultaneous variation experiment.

REFERENCES

Adelson, E. (2000). Lightness Perception and Lightness Illusions. In M. S. Gazzaniga, *The New Cognitive Neurosciences* (pp. 339-351). Cambridge, MA: The MIT Press.

American Society for Testing and Materials. (2017). *Standard test method for luminous reflectance factor of acoustical materials by use of integrating-sphere reflectometers.*

Arend, L. E. (1993). How much does illuminant color affect unattributed colors? *Journal of the Optical Society of America, 10*(10), 2134-2147.

Arend, L. E., & Goldstein, R. (1990). Lightness and brightness over spatial illumination gradients. *Journal of the Optical Society of America A, 7*(10), 1929-1936.

Arend, L. E., & Robert, G. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America A, 4*(12), 2281-2285.

Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 1. Illuminance variation. *Perception & Psychophysics, 54*(4), 446-456.

Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 2. Reflectance variation. *Perception & Psychophysics, 54*(4), 457-468.

Aston, S., Radonjic, A., Brainard, D. H., & Hurlbert, A. C. (2019). Illumination discrimination for chromatically biased illuminations: Implications for color constancy. *Journal of Vision, 19*(3), 1-15.

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

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

Brainard, D. H., & Radonjic, A. (2004). Color constancy. *The visual neurosciences., 1*, 948-961.

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

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. In H. J, *Encyclopedia of Imaging Science and Technology.* Wiley-Interscience.

Bramwell, D. I., & Hurlbert, A. C. (1996). Measurements of colour constancy by using a forced-choice matching technique. *Perception, 25*(2), 229-241.

Craven, B. J., & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research, 32*(7), 1359-1366.

Delahunt, P. B., & Brainard, D. H. (2004). Does human color constancy incorporate the statistical regularity of natural daylight? *Journal of Vision, 4*(2), 57-81.

Ennis, R., & Doerschner, K. (2019). Disentangling simultaneous changes of surface and illumination. *Vision Research, 158*, 173-188.

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

Gilchrist, A. (2006). *Seeing black and white.* Oxford University Press.

Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision, 7*(4), 1-15.

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

Jakob, W. (2010). Mitsuba Renderer.

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

709 Linhares, J. M., Pinto, P. D., & Nascimento, S. M. (2008). The number of discernible colors in
710   natural scenes. *Journal of the Optical Society of America, 25*(12), 2918-2924.

711 Luo, M. R., Clarke, A. A., Rhodes, P. A., Schappo, A., Scrivener, S. A., & Tait, C. J. (1991).
712   Quantifying colour appearance. Part I. LUTCHI colour appearance data. *Color Research*
713   *& Application, 16*(3), 166-180.

714 Maria, O., Hansen, T., & Gegenfurtner, a. K. (2009). Categorical color constancy for simulated
715   surfaces. *Journal of Vision, 9*(12), 1-6.

716 Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color
717   constancy for real surfaces. *Journal of Vision, 10*(9), 1-16.

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

721 Reeves, A. J., Amano, K., & Foster, D. H. (2008). Color constancy: phenomenal or projective?
722   *Perception & psychophysics., 70*(2), 219-228.

723 Rutherford, M. D., & Brainard, D. H. (2002). Lightness constancy: A direct test of the
724   illumination-estimation hypothesis. *Psychological Science, 13*(2), 142-149.

725 Schultz, S., Doerschner, K., & Maloney, L. T. (2006). Color constancy and hue scaling. *Journal*
726   *of Vision, 6*(10), 1-10.

727 Singh, V., Burge, J., & Brainard, D. H. (2022). Equivalent noise characterization of human
728   lightness constancy. *Journal of Vision, 22*(5), 1-26.

729 Singh, V., Cottaris, N. P., Heasly, B. S., Brainard, D. H., & Burge, J. (2018). Computational
730   luminance constancy from naturalistic images. *Journal of Vision, 18*(3), 1-19.

731 Smithson, H., & Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and
732   levels of reference. *Journal of Vision, 4*(9), 1-3.

733 Speigle, J. M., & Brainard, D. H. (1996). Is color constancy task independent? *Color and*
734   *Imaging Conference. 1*, pp. 167-172. Society for Imaging Science and Technology.

735 Troost, J. M., & De Weert, C. M. (1991). Naming versus matching in color constancy.
736   *Perception & psychophysics, 50*, 591-602.

737 Uchikawa, K., Uchikawa, H., & Boynton, R. M. (1989). Partial color constancy of isolated
738   surface colors examined by a color-naming method. *Perception, 18*(1), 83-91.

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

741