

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:** The lightness of an object is an intrinsic property that depends on its surface reflectance
7 spectrum. The visual system estimates an object's lightness from the light reflected off its surface. The
8 light reflected also depends on object extrinsic properties of the scene. For stable perception, the visual
9 system needs to discount variations due to extrinsic properties. We characterize this perceptual stability
10 for variation in two spectral properties of the scene: the reflectance spectra of background objects and the
11 intensity of light sources. We use a two-alternative forced-choice task to measure human observers'
12 thresholds of discriminating computer-generated images of 3D scenes based on the lightness of a
13 spherical target object in the scene. We measured how the discrimination thresholds changed as we varied
14 the reflectance spectra of the objects and the intensity of the light sources in the scene, both individually
15 and simultaneously. For small amounts of extrinsic variations, the thresholds of discrimination remained
16 constant indicating that the thresholds were dominated by observers' intrinsic representation of lightness.
17 As extrinsic variation increased, it started affecting observers' lightness judgment and the thresholds
18 increased. We estimated that the effects of extrinsic variations were comparable to observers' intrinsic
19 variation in the representation of object lightness. Moreover, for simultaneous variation of these spectral
20 properties, the increase in threshold square compared to no variation condition was a linear sum of the
21 corresponding increase in threshold squares for the individual properties, indicating that the variation
22 from these independent sources combines linearly.

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

24 **PRECIS:** We measure human lightness discrimination thresholds as a function of the amount of variation
25 in object-extrinsic spectral properties of visual scenes. We show that the visual system largely
26 compensates for such variations and that the effect of variation in independent properties combines
27 linearly.

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

38 The perceptual correlate of the diffuse spectral reflectance of an object is its perceived color. For
39 achromatic objects, the analogous perceptual quantity is object lightness. The human visual system is
40 known to provide a relatively stable representation of the color/lightness of an object despite variability in
41 the proximal signal due to changes in the light source, the surface reflectance of objects in the scene, and
42 the geometry and other properties of the scene (Foster, 2011; Brainard & Radonjic, 2004). The degree to
43 which such stability can be achieved is termed as color/lightness constancy (Adelson, 2000; Gilchrist,
44 2006). Human color/lightness constancy has been measured using appearance-based approaches and
45 discrimination-based approaches (Olkonen & Ekroll, 2016). Appearance based approaches involve tasks
46 in which the observer makes judgement about the appearance of stimuli. This approach includes methods
47 such as color matching, color naming, scaling, and nulling (Foster, 2003). In color matching, observers
48 adjust a test stimulus to match a standard stimulus. Color matching experiments show varying degrees of
49 constancy with constancy measured between 15%-90% under conditions such as changes of illumination
50 (Arend & Goldstein, 1987; Arend & Spehar, 1993), reflectance (Arend & Spehar, 1993; Patel,
51 Munasinghe, & Murray, 2018), illumination gradients (Arend & Goldstein, 1990; Brainard, Brunt, &
52 Speigle, 1997), and illumination and simulated reflectance (Rutherford & Brainard, 2002). Color naming
53 is a more direct and arguably natural method to measure color constancy where observers are asked to
54 categorize stimuli based on their hue, saturation, and lightness (Troost & De Weert, 1991). This method
55 has been used with real (Uchikawa, Uchikawa, & Boynton, 1989; Olkkonen, Witzel, Hansen, &
56 Gegenfurtner, 2010) and simulated stimuli (Olkonen, Hansen, & Gegenfurtner, 2009) to measure
57 constancy. Color naming methods have the limitation that there are vast number of possible discernible
58 colors (Linhares, Pinto, & Nascimento, 2008), but there is a limit of the gamut that can be displayed.
59 Typically, observers are asked to name from a small set of colors (Speigle & Brainard, 1996; Smithson &
60 Zaidi, 2004; Hansen, Walter, & Gegenfurtner, 2007) which might provide an overestimate of the
61 measured constancy. In color scaling methods, observers view a stimulus and provide a rating on a scale
62 for a set of colors, thus allowing for a finer level of comparison for measuring constancy (Luo, et al.,
63 1991; Schultz, Doerschner, & Maloney, 2006). Scaling methods can also be used to measure changes in
64 stimuli, where observers provide a rating of the change between stimuli (Ennis & Doerschner, 2019).
65 Nulling or achromatic adjustment methods involve changing a test stimulus such that it appears
66 achromatic (Arend, 1993; Brainard, 1998; Delahunt & Brainard, 2004). This method has the limitation
67 that it provides data only for achromatic/gray stimuli and additional assumptions about the observers'
68 criterion needs to be made for color appearances (Speigle & Brainard, 1996).

69 Discrimination-based approaches provide an objective method to measure color constancy (Bramwell &
70 Hurlbert, 1996; Reeves, Amano, & Foster, 2008). In these experiments, observers discriminate stimuli as
71 to whether they are the same or different from each other. The stimuli are varied in some relevant
72 parameter space to measure the threshold for discriminating changes in the parameter (Craven & Foster,
73 1992; Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert, 2014; Aston, Radonjic, Brainard, &
74 Hurlbert, 2019). Recently, Singh et. al (Singh, Burge, & Brainard, 2022) developed an equivalent noise
75 paradigm that relates thresholds of discrimination to the variability in observers' intrinsic representation
76 of object properties (e.g., its lightness) and the variability due to object extrinsic properties of the scene.

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

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

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

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

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

113 **2 EXPERIMENTAL METHODS**

114 **Overview**

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

121 Similar to the previous work, we used a two-alternative forced-choice (2AFC) procedure to measure
122 thresholds (Figure 1). On each trial of the experiment, observers viewed pairs of computer-generated 3D
123 scenes displayed on a color calibrated monitor. These pairs consisted of a standard image and a
124 comparison image. The images were presented sequentially for 250ms, with a 250ms inter-stimulus
125 interval between them. Both images contained a centrally located achromatic sphere as the target object.
126 Observers were required to indicate the image on which the target object was lighter. Between trials, we
127 manipulated the luminous reflectance factor (LRF) of the target object in the comparison image. The LRF
128 is the ratio of the luminance of a surface under a reference illuminant (here, the CIE D65 reference
129 illuminant) to the luminance of the reference illuminant itself (American Society for Testing and
130 Materials, 2017). The order of the standard and the comparison image was chosen in a pseudorandom
131 order. We recorded the proportion of times the observer chose the comparison image to have the lighter
132 target object at 11 values of the target object LRF. The psychometric function of one observer is shown in
133 Figure 2. We fit the proportion-comparison-chosen data with a cumulative normal function to extract the
134 thresholds, which was defined as the difference between the LRF of the target object at proportion
135 comparison chosen 0.76 and 0.50 (i.e., d' = 1.0 in a two-interval task).

136 We measured the effect of variation in two types of object-extrinsic scene properties on human lightness
137 discrimination thresholds: variation in the reflectance spectra of the background objects in the scene and
138 variation in the intensity of the light sources in the scene. We performed three experiments. These
139 experiments were preregistered (see below Preregistration).

140 (1) Background reflectance spectra variation (preregistered as Experiment 6): In this experiment, we
141 measured human lightness discrimination thresholds as a function of the amount of variation in the
142 background objects while the spectra of the light sources were kept fixed.

143 (2) Light source intensity variation (preregistered as Experiment 7): In this experiment, we measured
144 lightness discrimination thresholds as a function of the amount of variation in the intensity of the light
145 sources while the background was fixed.

146 (3) Simultaneous variation (preregistered as Experiment 8): In this experiment, we measured lightness
147 discrimination thresholds as both the background object reflectance spectra and the light source intensity
148 varied simultaneously.

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

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

161 The subsections below provide additional methodological detail.

162 **Preregistration**

163 We preregistered the experiments performed in this work before collecting the data. The preregistration
164 documents are available at: <https://osf.io/7tgy8/>.¹ The documentation includes information about the
165 experimental design as well as the procedure to estimate thresholds from the collected data.

166 The experiments were preregistered as Experiment 6 (referred here as Background reflectance spectra
167 variation), Experiment 7 (referred here as Light source intensity variation), and Experiment 8 (referred
168 here as Simultaneous variation). Experiment 6 was a replication of previous work (preregistered as
169 Experiment 3; Singh, Burge, & Brainard, 2022) with additional conditions in which the background
170 objects were achromatic and varied only in their lightness. While the stimuli were different for the three
171 experiments (preregistered Experiments 6, 7, and 8), the experimental method to measure lightness
172 discrimination thresholds were the same.

173 The preregistration documents mentioned that the experiments aimed at characterizing the dependence of
174 human lightness discrimination thresholds on the amount of variation in the background reflectance and
175 the intensity of the light source in the scene. The method of estimating discrimination thresholds was
176 described in the document. We predicted that the thresholds would increase with increase in the amount
177 of variation. For background variation, we predicted that the thresholds of achromatic variation would be
178 lower than chromatic variation. We also predicted that the increase in thresholds could be captured by an
179 equivalent noise model (Singh, Burge, & Brainard, 2022). Additionally, we predicted that the threshold
180 for simultaneous variation would be higher than the thresholds for individual variations.

181 Reflectance and Illumination Spectra

182 We used a statistical model of natural reflectance dataset to generate reflectance spectra of background
183 objects (Singh, Cottaris, Heasly, Brainard, & Burge, 2018; Singh, Burge, & Brainard, 2022). We
184 combined two datasets of surface reflectance measurements (Vrhel, Gershon, & Iwan, 1994; Kelly,
185 Gibson, & Nickerson, 1943). These datasets contain 632 surface reflectance measurements. We mean
186 centered the dataset by subtracting out the mean surface reflectance over the 632 measurements. Then we
187 used principal component analysis (PCA) to obtain the projection of the mean centered dataset along the
188 eigenvectors associated with the six largest eigenvalues. These eigenvalues captured more than 99.5% of
189 the variance (Singh, Cottaris, Heasly, Brainard, & Burge, 2018). The empirical distribution of the
190 projection weights thus obtained was approximated with a multivariate normal distribution. To get the
191 projection weights of random samples of reflectance spectra, pseudorandom samples were generated from
192 this multivariate normal distribution. Reflectance spectra were constructed by using these projection
193 weights along with the eigenvectors and adding the mean of the surface reflectance dataset. A physical
194 realizability condition was imposed on these spectra by ensuring that the reflectance at each wavelength
195 was between 0 and 1. If a reflectance spectrum did not meet this criterion, it was discarded.

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

199 To control the amount of variation in the reflectance spectra, the covariance matrix of the multivariate
200 normal distribution was multiplied by a covariance scalar (σ^2). A covariance scalar of 0 indicates that
201 there is no variation in the reflectance spectra of the background objects. On the other hand, a covariance

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.

202 scalar of 1 corresponds to the range of reflectance variation observed in the combined natural reflectance
203 datasets.

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

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

213 Background object reflectance variation (Experiment 6): In this experiment, we generated images for nine
214 conditions. Six of these conditions were for chromatic variation at six logarithmically spaced values of the
215 covariance scalar (σ^2): [0, 0.01, 0.03, 0.1, 0.3, 1.0] (same as Singh, Burge, & Brainard, 2022). Three
216 conditions were for achromatic variation at covariance scalar (σ^2): [0.03, 0.3, 1.0]. The power spectrum
217 of the light source was the same for all images. The power spectrum multiplication scalar was assigned an
218 arbitrary value of 5. Figure 3 shows five typical images for the nine conditions.

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

224 Simultaneous variation (Experiment 8): In this experiment we studied six conditions. These were: no
225 variation ($\sigma^2 = 0, \delta = 0$), chromatic background variation (covariance scalar = 1, $\delta = 0$), achromatic
226 background variation ($\sigma^2 = 1, \delta = 0$), light source intensity variation ($\sigma^2 = 0, \delta = 0.3$), simultaneous
227 variation chromatic background ($\sigma^2 = 1, \delta = 0.3$) and simultaneous variation achromatic background (σ^2
228 = 1, $\delta = 0.3$). Figure 5 shows five typical images for these six conditions.

229 **Stimulus Design**

230 The images used in this work were generated using the software Virtual World Color Constancy (VWCC)
231 (github.com/BrainardLab/VirtualWorldColorConstancy) described in (Singh, Cottaris, Heasly, Brainard,
232 & Burge, 2018). The initial step to generate an image involves constructing a 3D model that serves as the
233 base scene. Then, based on the specific experimental condition, we assign reflectance spectra and spectral
234 power distribution functions to the objects and light sources within the base scene. All light sources
235 within a given scene were assigned identical spectral power distribution functions. Subsequently, we
236 utilize Mitsuba, an physically-realistic open-source rendering system (mitsuba-renderer.org; Jakob, 2010)
237 to produce a 2D multispectral image of the scene at 31 equally spaced wavelengths between 400nm and
238 700nm. The images had a camera field of view of 17° and a resolution of 320-pixel by 240-pixels and
239 were centered at the target object. To display the images on the monitor, a 201-pixel by 201-pixel part
240 centered at the target object was cropped out of the images. Monitor calibration data was used to convert
241 the multispectral images to gamma corrected RGB images as described in (Singh, Burge, & Brainard,
242 2022). Gamma-corrected RGB images were presented on the calibrated 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 ($\sigma^2 = 0.00$, $\delta = 0.00$) 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 scene geometry
249 remained fixed during the experiment and the images did not include secondary reflections.

250 The standard image, when presented on the experimental monitor, had an average luminance of 87.1
251 cd/m² for the condition ($\sigma^2 = 0.00$, $\delta = 0.00$). The average luminance of the target object for the 11 LRF
252 levels were [120.9, 122.3, 123.8, 125.2, 126.5, 127.9, 129.2, 130.5, 131.9, 133.1, 134.4] cd/m².

253 For the ($\sigma^2 = 1.00$, $\delta = 0.30$) condition, the average luminance of the standard image was 87.8 cd/m². The
254 average luminance of the target object for the 11 LRF levels were [117.7, 119.4, 119.4, 122.3, 123.7,
255 123.8, 127.8, 126.9, 127.7, 129.1, 129.0] cd/m².

256 **Experimental Structure:**

257 In this study, a trial is defined as the display of a standard and a comparison image on the monitor and the
258 recording of the observer's response. An interval is defined as the presentation of either the standard
259 image or the comparison image within a trial. A block consists of recording 330 trials for one condition,
260 30 trials each at 11 comparison image target LRF levels. A permutation consists of recording one block of
261 data for each condition in an experiment. We recorded three permutations for each observer in each
262 experiment. Each permutation had 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 blocks (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 (nearly two to five minutes) between blocks. The observers were informed that they could stop the
271 experiment at any time. If the observer terminated a block, the data was not recorded. No observer
272 terminated a block 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 The experimental procedure was explained to each observer at the beginning of the practice session. The
278 experimenter then obtained the consent for the experiment. Vision tests were performed on the observer
279 to ensure normal visual acuity and normal color vision. After this, the observer went to the experimental
280 room where they were familiarized with the experimental set-up by performing a familiarization block of
281 40 trials. Then the observers were dark adapted by sitting in the dark for about 5 minutes. Then the data
282 for the three blocks of the practice session was recorded. At the end of the practice session, the observers
283 were informed if they could 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.

288 **Observer Recruitment and Exclusion**

289 The study recruited observers from North Carolina Agricultural and Technical State University and the
290 local Greensboro community. The participants were compensated for their time. The observers underwent
291 a screening process to meet criteria such as a normal visual acuity of 20/40 or better (with corrective
292 eyewear if necessary) and normal color vision, which was assessed using pseudo-isochromatic plates
293 (Ishihara, 1977). The preregistration documents outlined these exclusion criteria (see Methods:
294 Preregistration).

295 We further conducted a practice session to identify observers who could reliably perform the
296 psychophysical task. In the practice session, the observers' performed three blocks of the experiment for
297 the no variation condition ($\sigma^2 = 0.00$, $\delta = 0.00$) and the threshold was calculated for these three blocks.
298 If the mean threshold of the observer for the last two blocks in the practice session was larger than 0.030
299 ($\log T^2$, -3.2), the observer was discontinued. The preregistration document specified this exclusion
300 criteria (See Methods: Preregistration).

301 If the observers met these criteria, they were continued with the rest of the experiment.

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

304 **Observer Information**

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

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

328 their thresholds for two conditions did not fit the expected pattern. We removed their data from the
329 analysis presented in this work. Their data and thresholds are provided as supplementary information (SI).

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

346 **Apparatus**

347 The experiments were performed in a dark room and the stimuli were presented on a color calibrated
348 LCD monitor (27-in. NEC MultiSync EA271U; NEC Display Solutions). The monitor operated at a pixel
349 resolution of 1920 x 1080, with a refresh rate of 60Hz, and 8-bit resolution for each RGB channel. The
350 setup utilized an Apple Macintosh computer with an Intel Core i7 processor. The experimental programs
351 were written in MATLAB (MathWorks; Natick, MA). The programs utilized the Psychophysics Toolbox
352 (<http://psychtoolbox.org>) and mgl (<http://justingardner.net/doku.php/mgl/overview>) libraries. A Logitech
353 F310 gamepad controller was used to collect observers' response.

354 The distance between the observers' eyes and the monitor was set at 75cm. To ensure stability and proper
355 alignment, a chin cup and forehead rest (Headspot, UHCOTech, Houston, TX) were used to stabilize the
356 observers' head position. The observers' eyes were centered both horizontally and vertically in relation to
357 the display.

358 **Monitor Calibration**

359 The monitor was calibrated using a spectroradiometer (PhotoResearch PR655) as described in (Singh,
360 Burge, & Brainard, 2022). The monitor was calibrated before starting each experiment. Once calibrated,
361 the same settings were used till data for all observers for that experiment was collected. The monitor was
362 then recalibrated for the next experiment. Data was collected in the sequence Experiment 6, Experiment
363 7, and Experiment 8.

364 **Ethics Statement**

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

368 **Code and Data Availability**

369 The data for each experiment and observer is provided as supplementary information (SI). The SI is
370 available at: <https://github.com/vijaysoophie/SimultaneousVariationPaper>. The SI contains the proportion
371 comparison chosen data as well as the thresholds for the 3 experimental blocks of each condition, for each
372 experiment and observer. The MATLAB scripts to generate Figures 2, 6 – 12, supplementary Figures S1-
373 S7, and the scripts to obtain thresholds of the linear receptive field formulation of the model are also
374 provided in the SI.

375 Linear Receptive Field Model

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

$$388 \quad T = \sqrt{\sigma_{ri}^2 + \sigma^2 \sigma_{eo}^2} \quad (1)$$

389 where σ^2 is the covariance scalar (see (Singh, Burge, & Brainard, 2022) for details). The variance of the
390 extrinsic properties (σ_{eo}^2) resulting from the variation in the image can be computed as $R^T \Sigma_{eo} R$, where
391 Σ_{eo} is the covariance matrix of the variation in the images. We calculate the pixel-by-pixel covariance
392 matrix of the image database for each condition and use the receptive field vector R to estimate the
393 quantity $R^T \Sigma_{eo} R$.

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

400 Light source intensity variation: We fit a functional form similar to Eq. (1) to the thresholds of light
401 source intensity variation experiment (Experiment 7) where we replaced covariance scalar σ^2 by the
402 range parameter δ . We chose the value of the Gaussian noise variance ($\sigma_{ri,L}^2$) and receptive field surround
403 sensitivity ($v_{s,L}$) to minimize the mean square difference between the observer and model thresholds
404 measured at the seven values of the range parameter. $\sigma_{ri,L}^2$ provided the estimate of the observer's intrinsic
405 noise.

406 In natural viewing conditions, the intensity of the light source varies over several orders of magnitude
407 (Singh, Cottaris, Heasly, Brainard, & Burge, 2018). This range of variation cannot be captured in the
408 experiment due to the limitations of the monitor. To estimate the extrinsic noise, we used the best fit

409 surround sensitivity ($v_{s,L}$) and the sample covariance matrices to calculate the quantity $R^T \Sigma_{e0} R$ as a
410 function of the range parameter δ . We fit the resulting values with an exponential function (see Figure
411 S1). The extrinsic noise ($\sigma_{e0,L}^2$) was chosen as the value of the exponential fit at $\delta = 1$. The parameter
412 ($\sigma_{e0,L}^2$) could be used to estimate the extrinsic noise for more naturalistic variations using the exponential
413 fit.

414 **3 RESULTS**

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

416 We measured lightness discrimination thresholds of human observers for two types of variation in the
417 reflectance spectra of background objects in the scene: chromatic variation and achromatic variation. In
418 chromatic variation, the reflectance spectra could take any shape and thus the background objects varied
419 in their chromaticity and luminance. In achromatic variation, each spectrum had the same reflectance at
420 all wavelengths, and thus the spectra varied only in their overall luminance and the objects were gray. The
421 reflectance spectra were sampled from a multivariate normal distribution which was a statistical model of
422 natural surface reflectance spectra. The amount of variation in the reflectance spectra was controlled by
423 multiplying the covariance matrix by a covariance scalar (σ^2). We measured discrimination thresholds of
424 six human observers at six values of the covariance scalar for chromatic variation and three values of
425 covariance scalar for achromatic variation. For each of the nine conditions and for each observer,
426 discrimination thresholds were measured three times in three separate blocks. The psychometric functions
427 for these nine conditions are shown for one observer in Figure 6. See Figure S2 for the psychometric
428 functions of all six observers. We notice that as the covariance scalar increases, the slope of the
429 psychometric functions decreases, corresponding to an increase in discrimination thresholds. The
430 thresholds for chromatic and achromatic variation are comparable.

431 Figures 7 shows the change in discrimination thresholds as function of the variance in the reflectance of
432 the background objects. Here we plot the mean log threshold squared (averaged across observers, $N = 6$)
433 as a function of the log of the covariance scalar. The thresholds and standard error of the mean (SEM)
434 from Figure 7 are listed in Table S1. We observe that the thresholds are nearly constant at small values of
435 the covariance scalar. Once the value of the covariance scalar starts increasing, log threshold squared
436 increases. Further, the thresholds are comparable for chromatic and achromatic variation. The p-values of
437 the hypothesis that the mean thresholds for chromatic and achromatic variations are equal are 0.72, 0.57,
438 and 0.16 for covariance scalar 0.03, 0.30, and 1.00 respectively, indicating that the differences in the
439 mean thresholds are not statistically significant. Figure S3 shows the discrimination thresholds measured
440 in preregistered Experiment 6 and previously reported thresholds in (Singh, Burge, & Brainard, 2022).
441 The measured thresholds are consistent between the two experiments.

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

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

452 We measured lightness discrimination thresholds of human observers as we varied the intensity of light
453 sources in the scene. The shape of the spectrum of the light sources was fixed to be standard daylight
454 spectrum D65. We normalized the spectrum by its mean over wavelengths. The intensity was varied by
455 multiplying the normalized spectrum by a scalar sampled from a log-uniform distribution in the range [1-
456 δ , $1 + \delta$]. The reflectance spectra of the background objects were fixed. We measured lightness
457 discrimination thresholds for seven values of the range parameter δ for five human observers. The
458 psychometric function of one of the observers for these seven conditions are shown in Figure 8. Figure S4
459 shows the psychometric functions of all observers. Figure 9 shows how the thresholds change as the
460 amount of variation in the light source intensity increases. The data is averaged over five observers (also
461 see Figure S5). Table S2 lists the mean thresholds and the SEM measured in this experiment. Similar to
462 the trend for reflectance spectra variation, lightness discrimination thresholds remain constant for small
463 values of the range parameter and then log threshold squared increases with increase in the range
464 parameter. A fit of the mean squared threshold with the linear receptive field model gives the value of
465 internal noise as $\sigma_{ri,L}^2 = 0.028$. This compares well with the internal noise obtained from the background
466 reflectance spectra variation experiment ($\sigma_{ri,B}^2 = 0.026$). The variance of the extrinsic variability
467 estimated at range parameter $\delta = 1.00$ is $\sigma_{eo,L}^2 = 0.052$. The equivalent noise level, the ratio of external
468 variation to intrinsic noise is ~ 1.8 . This indicates that the variation in the lightness representation induced
469 by the variation in light source intensity is close to the internal variation of that representation at these
470 levels. In natural conditions, where light source intensity varies over several order of magnitudes, the
471 extrinsic noise can be estimated by generating images at the level of variation and using the LINRF
472 model.

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

474 We measured lightness discrimination thresholds for simultaneous variation in the reflectance spectra of
475 background objects and the intensity of the light sources in the scene. In this experiment, we studied six
476 conditions: no variation, variation in the reflectance spectra of background objects with fixed spectrum of
477 the light sources for achromatic and chromatic backgrounds, variation in intensity of light source with
478 fixed background, and simultaneous variation in the intensity of light source and reflectance spectra of
479 background object for chromatic and achromatic backgrounds. We measured lightness discrimination
480 thresholds of six human observers for these six conditions. The psychometric function of one of the
481 observers is shown in Figure 10. Figure S6 shows the psychometric functions of all observers. Figure 11
482 shows the mean log squared threshold of all six observers for these six conditions. Table S3 lists the mean
483 thresholds and SEM from Figure 11. The threshold for simultaneous variation of light source intensity
484 and reflectance spectra of background objects is higher than the condition with individual variations in
485 these properties. As observed earlier, the threshold for achromatic and chromatic conditions are
486 comparable. The p-value of the hypothesis that the mean thresholds for chromatic and achromatic
487 variations are equal is 0.19 for background variation condition and 0.44 for simultaneous variation
488 condition, indicating that the differences in the mean thresholds are not statistically significant.

489 Figure 11 also shows the squared thresholds of the LINRF model for the six conditions. We used the
490 intrinsic noise and the surround sensitivity (v_s) parameters of the background reflectance variation
491 experiment to estimate the threshold of the LINRF model for the no-variation condition, background
492 spectra variation conditions, and the simultaneous variation conditions (Experiment 6, Figure 7). For the
493 light intensity variation condition, we used the parameters of the light source intensity variation
494 experiment (Experiment 7, Figure 9). See Figure S7 for the predictions of the LINRF model with same set
495 of parameters for all six conditions.

496 We can use the linear receptive field model to compare the extrinsic variance of the simultaneous
497 variation condition to the variance of the individual variations. According to the linear receptive model,

498 the square of the threshold is proportional to the sum of the variance of observers' intrinsic noise and the
499 extrinsic variation in the stimuli (Eq. 1). The squared threshold at the no variation condition is equal to the
500 variance of the observers' intrinsic noise. In case of extrinsic variation, the increase in threshold square
501 compared to the no variation condition equals the variance of the extrinsic variation. When there are more
502 than one independent sources of extrinsic variation, the total variance of the simultaneous variation
503 should be the sum of the variance of the individual variations. This predicts that the increase in threshold
504 square for simultaneous variation condition should be equal to the sum of the corresponding increase for
505 the individual variation conditions.

506 Figure 12 shows the increase in mean squared threshold above the no variation condition. We compare
507 the mean square thresholds of the simultaneous variation condition with the sum of the mean square
508 thresholds of the individual conditions for chromatic and achromatic conditions. The increase in squared
509 threshold of the simultaneous variation condition is comparable to the sum of the increase in squared
510 threshold for the individual variations. The p-value of the hypothesis that the mean increase in squared
511 thresholds for simultaneous variation is equal to sum of the mean increase in the squared thresholds of
512 light intensity variation and background object reflectance variation are 0.86 and 0.80 for chromatic and
513 achromatic conditions respectively. The variance of the extrinsic noise calculated for the background
514 variation condition ($\sigma^2 = 1.00, \delta = 0.00$) is 0.0015 and the light intensity variation condition ($\sigma^2 =$
515 $0.00, \delta = 0.30$) is 0.0017. As expected, the variance of the simultaneous variation condition ($\sigma^2 = 1.00,$
516 $\delta = 0.30$), which is 0.0033, is comparable to the sum of individual variances (0.0032).

517 4 DISCUSSION

518 The visual system is known to maintain a stable representation of object lightness despite variations in the
519 proximal signal due to the scene. In this work we characterize the extent of such stability for variability in
520 the spectra of background objects and light sources in a scene. We measured human observers' threshold
521 of discriminating two objects based on their lightness as a function of amount of variation in these
522 spectral properties. We observe that for low levels of variability, the thresholds first remained constant,
523 showing that in this regime the performance was determined by observers' intrinsic noise. As the
524 variability increased, the effect of extrinsic variation started dominating the performance and the
525 discrimination thresholds increased. Using a model based on signal detection theory, we related the
526 thresholds in the low variability regime to the internal noise of the observer. The model also related the
527 increase in threshold to the amount of variability in the extrinsic property, thus providing a comparison of
528 the variance in the extrinsic property to the intrinsic noise. The effect of both types of extrinsic variation,
529 spectra of background objects and intensity of light sources, were comparable to the effect of intrinsic
530 noise, showing that the visual system provides a good degree of constancy to these variations. Further, for
531 simultaneous variation of these properties the effects added linearly, resulting in the variance of the
532 simultaneous variation to be equal to the sum of the variance of the individual variations.

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

538 **Visual system at threshold level:** For the spectral variabilities studied in this work, the variances in
539 observers' representation of lightness due to extrinsic variations are within a factor of two compared to
540 the variances in observers' intrinsic representation of lightness. If these variances were equal, one could
541 conclude that the visual system has fully compensated for the extrinsic variation. As the extrinsic
542 variances are larger than the variance of the intrinsic noise, the visual system has not fully compensated
543 the external variabilities. But since these variances are within a factor or two, it shows that the visual

544 system provides a large degree of stability in the perceptual representation of lightness and seems to work
545 at near threshold levels.

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

550 **5 ACKNOWLEDGEMENTS:** We thank Dr. David Brainard for his comments on the manuscript. This
551 work was supported by National Science Foundation award BCS 2054900.

552
553
554

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

555
556
557 **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 \pm 0.0012	0.0181 \pm 0.0001	0.0208 \pm 0.0014	0.0520 \pm 0.0114	0.0329 \pm 0.0061	0.0372 \pm 0.0008
$\delta = 0.05$	0.0228 \pm 0.0018	0.0229 \pm 0.0018	0.0207 \pm 0.0007	0.0580 \pm 0.0064	0.0346 \pm 0.0042	0.0364 \pm 0.0013
$\delta = 0.10$	0.0275 \pm 0.0024	0.0217 \pm 0.0009	0.0242 \pm 0.0040	0.0325 \pm 0.0022	0.0343 \pm 0.0013	0.0376 \pm 0.0072
$\delta = 0.15$	0.0316 \pm 0.0009	0.0238 \pm 0.0011	0.0323 \pm 0.0032	0.0333 \pm 0.0019	0.0345 \pm 0.0042	0.0326 \pm 0.0002
$\delta = 0.20$	0.0447 \pm 0.0100	0.0381 \pm 0.0046	0.0276 \pm 0.0016	0.0493 \pm 0.0120	0.0423 \pm 0.0050	0.0392 \pm 0.0034
$\delta = 0.25$	0.0433 \pm 0.0052	0.0393 \pm 0.0062	0.0308 \pm 0.0023	0.0461 \pm 0.0060	0.0532 \pm 0.0083	0.0387 \pm 0.0025
$\delta = 0.30$	0.0404 \pm 0.0018	0.0429 \pm 0.0033	0.0347 \pm 0.0014	0.0580 \pm 0.0061	0.0465 \pm 0.0047	0.0421 \pm 0.0042

558

559
560
561

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

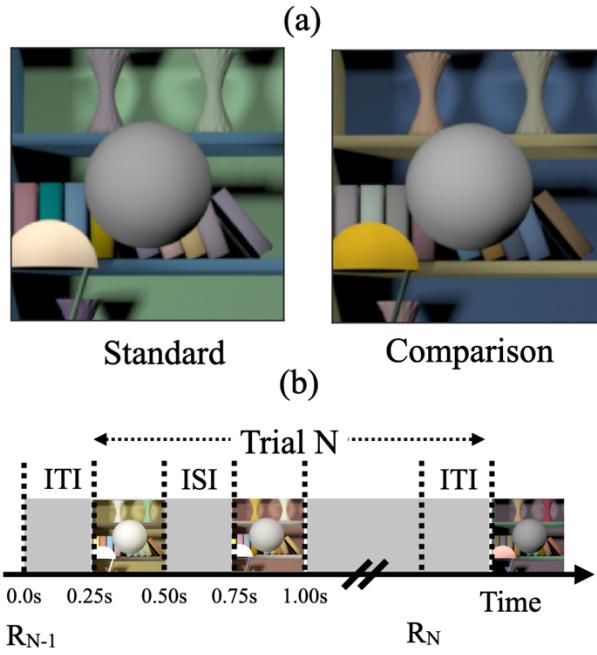
Condition	Observer					
	0003	Bagel	Content	Oven	Manos	Revival
No Variation	0.0261 \pm 0.0022	0.0227 \pm 0.0019	0.0246 \pm 0.0004	0.0383 \pm 0.0066	0.0258 \pm 0.0036	0.0366 \pm 0.0085
Background Variation Chromatic	0.0414 \pm 0.0036	0.0340 \pm 0.0058	0.0392 \pm 0.0083	0.0498 \pm 0.0050	0.0306 \pm 0.0013	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

562

706 REFERENCES

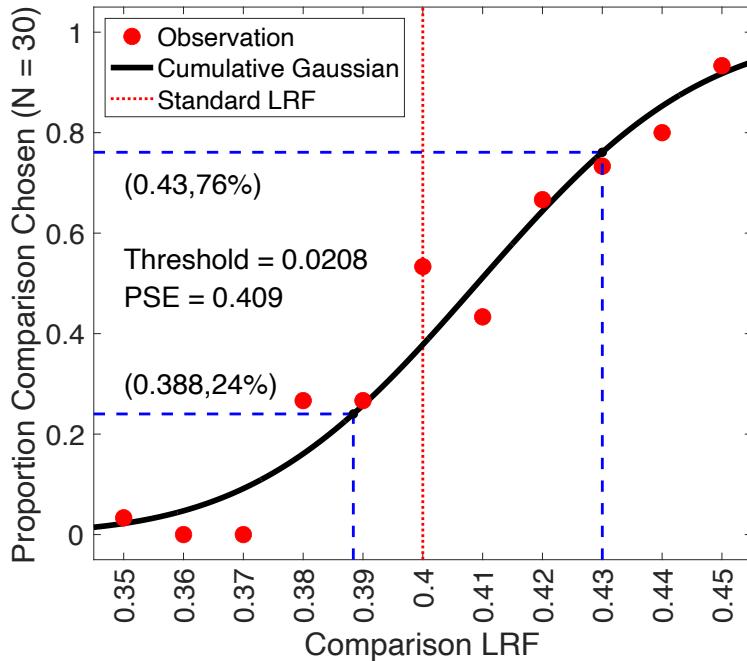
- 707
708 Adelson, E. (2000). Lightness Perception and Lightness Illusions. In M. S. Gazzaniga, *The New*
709 *Cognitive Neurosciences* (pp. 339-351). Cambridge, MA: The MIT Press.
710 American Society for Testing and Materials. (2017). *Standard test method for luminous*
711 *reflectance factor of acoustical materials by use of integrating-sphere reflectometers.*
712 Arend, L. E. (1993). How much does illuminant color affect unattributed colors? *Journal of the*
713 *Optical Society of America*, 10(10), 2134-2147.
714 Arend, L. E., & Goldstein, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal*
715 *of the Optical Society of America A*, 4(12), 2281-2285.
716 Arend, L. E., & Goldstein, R. (1990). Lightness and brightness over spatial illumination
717 gradients. *Journal of the Optical Society of America A*, 7(10), 1929-1936.
718 Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 1. Illuminance
719 variation. *Perception & Psychophysics*, 54(4), 446-456.
720 Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 2. Reflectance
721 variation. *Perception & Psychophysics*, 54(4), 457-468.
722 Aston, S., Radonjic, A., Brainard, D. H., & Hurlbert, A. C. (2019). Illumination discrimination
723 for chromatically biased illuminations: Implications for color constancy. *Journal of*
724 *Vision*, 19(3), 1-15.
725 Brainard, D. H. (1998). Color constancy in the nearly natural image. 2. Achromatic loci. *Journal*
726 *of the Optical Society of America A*, 15(2), 307-325.
727 Brainard, D. H., & Radonjic, A. (2004). Color constancy. *The visual neurosciences.*, 1, 948-961.
728 Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural
729 image. Asymmetric matches. *Journal of the Optical Society of America A*, 14(9), 2091-
730 2110.
731 Bramwell, D. I., & Hurlbert, A. C. (1996). Measurements of colour constancy by using a forced-
732 choice matching technique. *Perception*, 25(2), 229-241.
733 Craven, B. J., & Foster, D. H. (1992). An operational approach to colour constancy. *Vision*
734 *Research*, 32(7), 1359-1366.
735 Delahunt, P. B., & Brainard, D. H. (2004). Does human color constancy incorporate the
736 statistical regularity of natural daylight? *Journal of Vision*, 4(2), 57-81.
737 Ennis, R., & Doerschner, K. (2019). Disentangling simultaneous changes of surface and
738 illumination. *Vision Research*, 158, 173-188.
739 Foster, D. H. (2003). Does colour constancy exist? *Trends in cognitive sciences*, 7(10), 439-443.
740 Foster, D. H. (2011). Color constancy. *Vision Research*, 51(7), 674-700.
741 Gilchrist, A. (2006). *Seeing black and white*. Oxford University Press.
742 Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on
743 color categories and color constancy. *Journal of Vision*, 7(4), 1-15.
744 Ishihara, S. (1977). *Tests for colour-blindness*. Tokyo: Kanehara Shuppon Company, Ltd.
745 Jakob, W. (2010). Mitsuba Renderer.
746 Kelly, K. L., Gibson, K. S., & Nickerson, D. (1943). Tristimulus specification of the Munsell
747 book of color from spectrophotometric measurements. *Journal of the Optical Society of*
748 *America*, 33(7), 355-376.
749 Linhares, J. M., Pinto, P. D., & Nascimento, S. M. (2008). The number of discernible colors in
750 natural scenes. *Journal of the Optical Society of America*, 25(12), 2918-2924.

- 751 Luo, M. R., Clarke, A. A., Rhodes, P. A., Schappo, A., Scrivener, S. A., & Tait, C. J. (1991).
752 Quantifying colour appearance. Part I. LUTCHI colour appearance data. *Color Research*
753 & Application, 16(3), 166-180.
- 754 Olkkonen, M., & Ekroll, V. (2016). Color constancy and contextual effects on color appearance.
755 In J. Kremers, R. Baraas, & N. Marshall, *Human color vision, Springer Series in Vision*
756 Research Vol. 5 (pp. 159-188). Springer, Cham.
- 757 Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2009). Categorical color constancy for
758 simulated surfaces. *Journal of Vision*, 9(12), 1-6.
- 759 Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color
760 constancy for real surfaces. *Journal of Vision*, 10(9), 1-16.
- 761 Patel, K. Y., Munasinghe, A. P., & Murray, R. F. (2018). Lightness matching and perceptual
762 similarity. *Journal of vision*, 18(5), 1-13.
- 763 Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. (2014). Chromatic
764 illumination discrimination ability reveals that human colour constancy is optimised for
765 blue daylight illuminations. *Plos One*, 9(2), e87989.
- 766 Reeves, A. J., Amano, K., & Foster, D. H. (2008). Color constancy: phenomenal or projective?
767 *Perception & psychophysics*, 70(2), 219-228.
- 768 Rutherford, M. D., & Brainard, D. H. (2002). Lightness constancy: A direct test of the
769 illumination-estimation hypothesis. *Psychological Science*, 13(2), 142-149.
- 770 Schultz, S., Doerschner, K., & Maloney, L. T. (2006). Color constancy and hue scaling. *Journal*
771 *of Vision*, 6(10), 1-10.
- 772 Singh, V., Burge, J., & Brainard, D. H. (2022). Equivalent noise characterization of human
773 lightness constancy. *Journal of Vision*, 22(5), 1-26.
- 774 Singh, V., Cottaris, N. P., Heasly, B. S., Brainard, D. H., & Burge, J. (2018). Computational
775 luminance constancy from naturalistic images. *Journal of Vision*, 18(3), 1-19.
- 776 Smithson, H., & Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and
777 levels of reference. *Journal of Vision*, 4(9), 1-3.
- 778 Speigle, J. M., & Brainard, D. H. (1996). Is color constancy task independent? *Color and*
779 *Imaging Conference*. 1, pp. 167-172. Society for Imaging Science and Technology.
- 780 Troost, J. M., & De Weert, C. M. (1991). Naming versus matching in color constancy.
781 *Perception & psychophysics*, 50, 591-602.
- 782 Uchikawa, K., Uchikawa, H., & Boynton, R. M. (1989). Partial color constancy of isolated
783 surface colors examined by a color-naming method. *Perception*, 18(1), 83-91.
- 784 Vrhel, M. J., Gershon, R., & Iwan, L. S. (1994). Measurement and analysis of object reflectance
785 spectra. *Color Research & Application*, 19(1), 4-9.
- 786



1

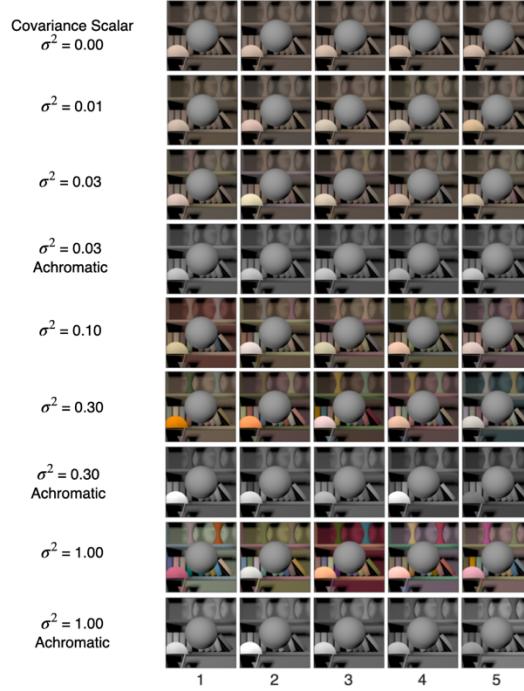
2 **Figure 1: (a) Psychophysical task.** (Adapted from Figure 1 in Singh, Burge, & Brainard, 2022) The
 3 psychophysical task involved comparing two images, a standard image and a comparison image, on each
 4 trial and selecting the image with the lighter target object. The target object was an achromatic sphere at
 5 the center of the image. The images were computer graphics renderings of 3D scenes. They were
 6 displayed on a color calibrated monitor. This panel shows examples of standard and comparison images.
 7 The reflectance spectrum of the target object was spectrally flat, and the target object appeared gray. The
 8 reflectance of the target object in the standard image was held fixed and it changed for the comparison
 9 image. In this panel, the target object in the comparison image is lighter. We measured the fraction of
 10 times the observers chose the target object in the comparison image to be lighter as a function of the
 11 lightness of the target object in the comparison image. Fraction comparison chosen data was used to
 12 determine lightness discrimination threshold (Figure 2). We studied how the lightness discrimination
 13 thresholds changed as the trial-to-trial variability in the reflectance spectra of the background objects and
 14 the intensity of the light sources increased. **(b) Trial sequence:** R_{N-1} indicates the recording of the
 15 observer's response for the (N-1)th trial. The Nth trial begins 250ms after the completion of the (N-1)th trial
 16 (Inter Trial Interval, ITI = 250ms). In the Nth trial, the standard and comparison images are presented for
 17 250ms each with a 250ms inter stimulus interval (ISI) in between the two images. The order of the
 18 standard and comparison images is chosen in pseudorandom order. The observer records their choice by
 19 pressing a button on a gamepad after both images have been presented and removed from the screen. The
 20 observers could take as long as they wish before making their choice. The recording of their choice is
 21 indicated by R_N in the panel. The next trial begins 250ms after the choice has been recorded.



1

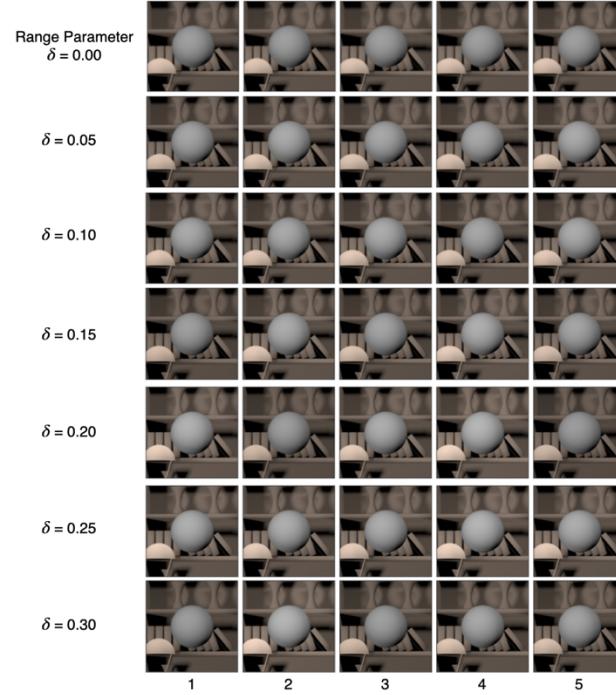
2 **Figure 2: Psychometric function:** We measured the proportion of times the observers selected the target
 3 in the comparison image to be lighter as function of the LRF (lightness reflectance factor) of the target
 4 object. We collected 30 responses for each of the 11 equally spaced values of the comparison image target
 5 object LRF, ranging from 0.35 to 0.45. The LRF of the target object in the standard image was 0.40. The
 6 LRF of the target object in the comparison image was selected in a pseudorandom order. To analyze the
 7 data, we used maximum likelihood methods to fit a cumulative normal function to the proportion
 8 comparison chosen data. We imposed constraints on the guess rate and lapse rate, requiring them to be
 9 equal and within the range of 0 to 0.05. The threshold was determined as the difference between the LRF
 10 values corresponding to a proportion comparison chosen of 0.76 and 0.50, obtained from the cumulative
 11 normal fit. The figure presented here illustrates the data for observer 0003 in the second block of the
 12 background reflectance variation experiment (previously registered as Experiment 6) for the no variation
 13 ($\sigma^2 = 0.00, \delta = 0.00$) condition. The discrimination threshold was measured to be 0.0208. The point of
 14 subjective equality (PSE), which corresponds to a proportion of 0.5 in the comparison task, was found to
 15 be 0.409. The lapse rate for this particular fit was determined to be 0.00.
 16

1



2 **Figure 3: Background object reflectance variation:** We studied two types of variations in the
 3 reflectance spectra of background objects in the scene: chromatic variation and achromatic variation. In
 4 chromatic variation, the reflectance spectra could take any shape, and the objects varied in their
 5 luminance and chromaticity. In achromatic variation, the reflectance spectra were spectrally flat, and the
 6 objects appeared gray and varied only in their luminance. The spectra were chosen from a multivariate
 7 normal distribution that modeled the statistics of natural reflectance spectra. The variation in the
 8 reflectance spectra was controlled by multiplying the covariance matrix of the distribution with a scalar.
 9 We generated images at six logarithmically spaced values of the covariance scalar for chromatic variation
 10 and at three values of the covariance scalar for achromatic variations. The figure shows five typical
 11 images for each of these nine conditions. For each condition we generated 1100 images, 100 images at 11
 12 linearly spaced value of target object LRF in the range [0.35, 0.45]. The target object in each image in the
 13 figure is at LRF = 0.4.

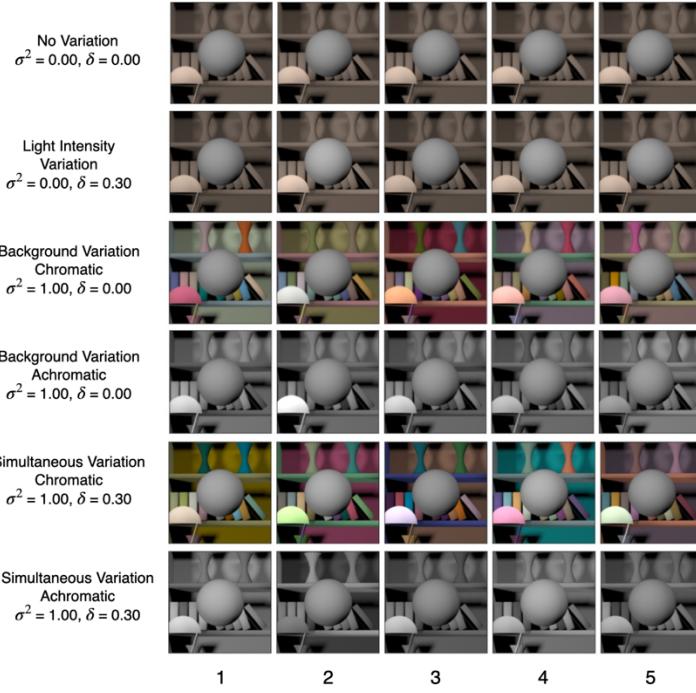
1

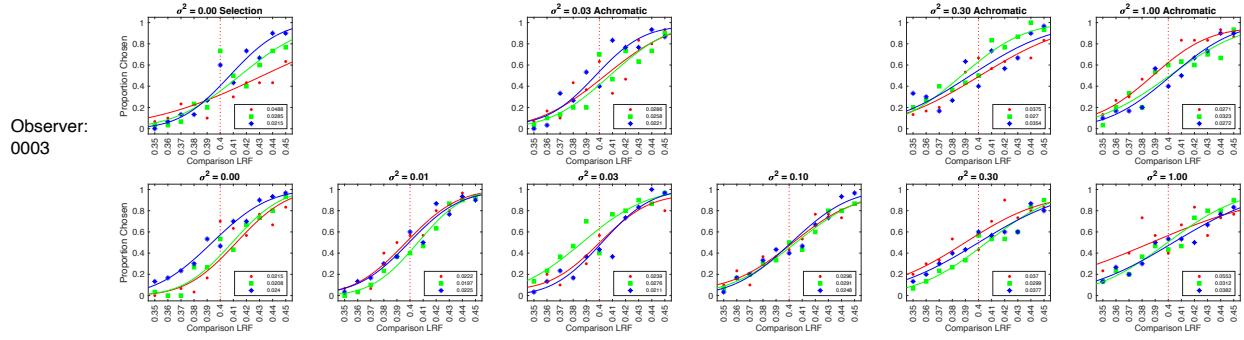


2 **Figure 4: Light intensity variation:** The shape of the power spectrum of the light sources in the scene
3 was chosen to be CIE reference illuminant D65. The intensity of the power spectrum was varied by
4 multiplying the normalized D65 spectrum with a scalar sampled from a log uniform distribution in the
5 range $[1 - \delta, 1 + \delta]$. The amount of variation was controlled by changing the value of the range parameter
6 δ . We generated images at seven linearly spaced values of the range parameter in the range $[0.00, 0.30]$.
7 For each value of the range parameter, we generated 1100 images, 100 images at each value of the target
8 object LRF in the range $[0.35, 0.45]$. The figure shows five sample images at each of the seven values of
9 the range parameter. The target object in each image in the figure has the same LRF of 0.40.

1

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

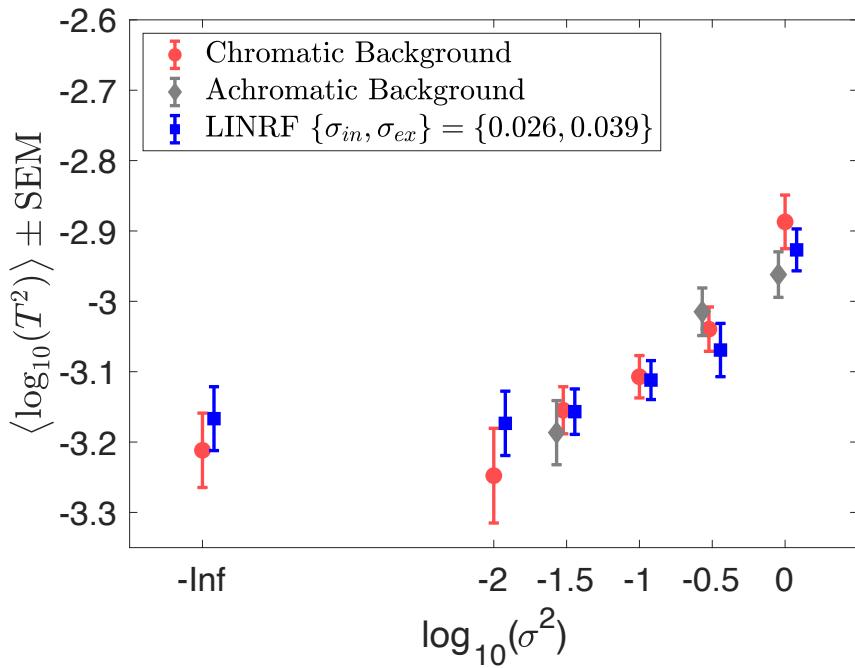




1

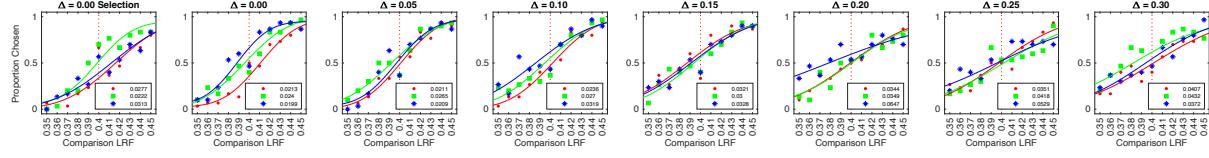
2 Figure 6: Psychometric functions for observer 0003 for background reflectance variation

3 **experiment:** We measured the proportion comparison chosen data for the nine conditions separately in
 4 three blocks for each observer. The figure shows the psychometric function for observer 0003. The
 5 psychometric functions for all six observers are shown in Figure S2. A cumulative normal function was
 6 fit to the data from each block to determine the discrimination threshold (see Figure 2). The legend
 7 provides the estimated lightness discrimination threshold for each block, obtained from the cumulative fit.
 8 The first panel in the top row shows the data and thresholds for the selection session. The selection
 9 session was a practice session in which the thresholds for the no variation condition was measured three
 10 times. An observer was selected for the experiment only if the average of their last two discrimination
 11 threshold measurements in the selection session was less than 0.30. The last three panels in the top row
 12 show the data for the three achromatic conditions. The bottom row shows the data for the chromatic
 13 variation conditions.

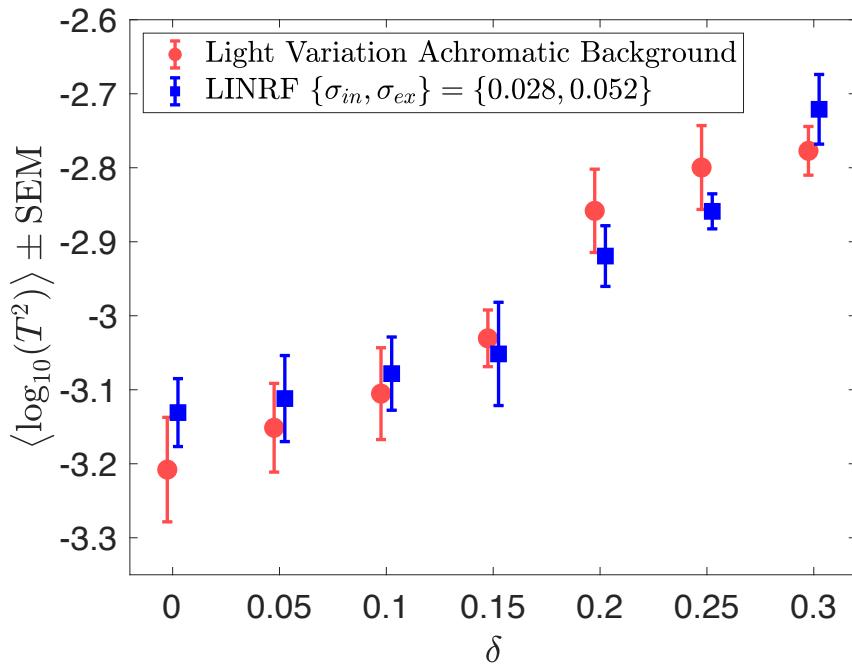


1

2 **Figure 7: Background variation increases lightness discrimination thresholds.** Mean (N = 6) log
 3 squared threshold vs log covariance scalar from human psychophysics for chromatic (red circles) and
 4 achromatic conditions (gray diamonds). The error bars represent +/- 1 SEM taken between observers. The
 5 threshold of the linear receptive field (LINRF) model was estimated by simulation for the six values of
 6 the covariance scalar (blue squares). The blue error bars show +/- 1 standard deviation estimated over 10
 7 independent simulations. The legend shows the parameters of the linear receptive field (LINRF) model
 8 fit. The data has been jittered for ease of viewing. A comparison of the thresholds with the previously
 9 published data in Singh, Burge, Brainard 2022 is shown in Figure S3.



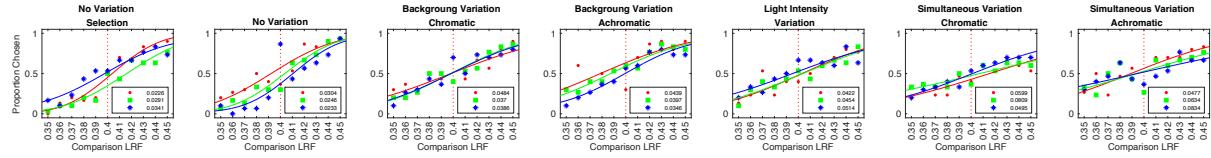
1
2 **Figure 8: Psychometric functions for observer 0003 for light intensity variation experiment:** Same
3 as Figure 6, but for the light intensity variation experiment. The figure shows the proportion comparison
4 chosen data for the selection session and the seven condition for observer 0003. The psychometric
5 functions for all observers are shown in Figure S4.



1

2 **Figure 9: Light source intensity variation increases lightness discrimination threshold.** Mean (N = 5)
3 log squared threshold vs range parameter from human psychophysics for the seven light source intensity
4 variation conditions (red circles). The error bars represent +/- 1 SEM taken between observers. The
5 threshold of the linear receptive field (LINRF) model was estimated by simulation for the seven values of
6 the range parameters (blue squares). The blue error bars show +/- 1 standard deviation estimated over 10
7 independent simulations. The parameters of the LINRF fit are provided in the legend. The data has been
8 jittered for ease of viewing. The data for all six observers is shown in Figure S5.

1



2 **Figure 10: Psychometric functions for observer 0003 for simultaneous variation experiment:** Same
3 as Figure 6 and 8, but for simultaneous variation experiment. The figure shows the proportion comparison
4 chosen data for the selection session and the six condition for observer 0003. The data for all observers
5 are shown in Figure S6.

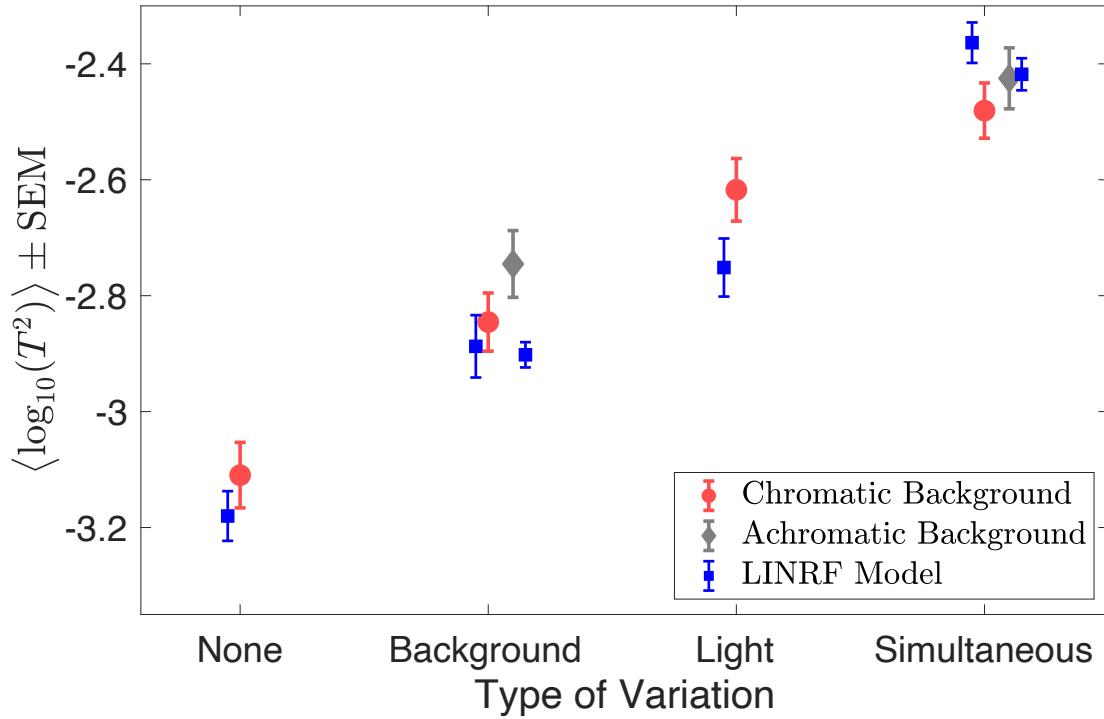
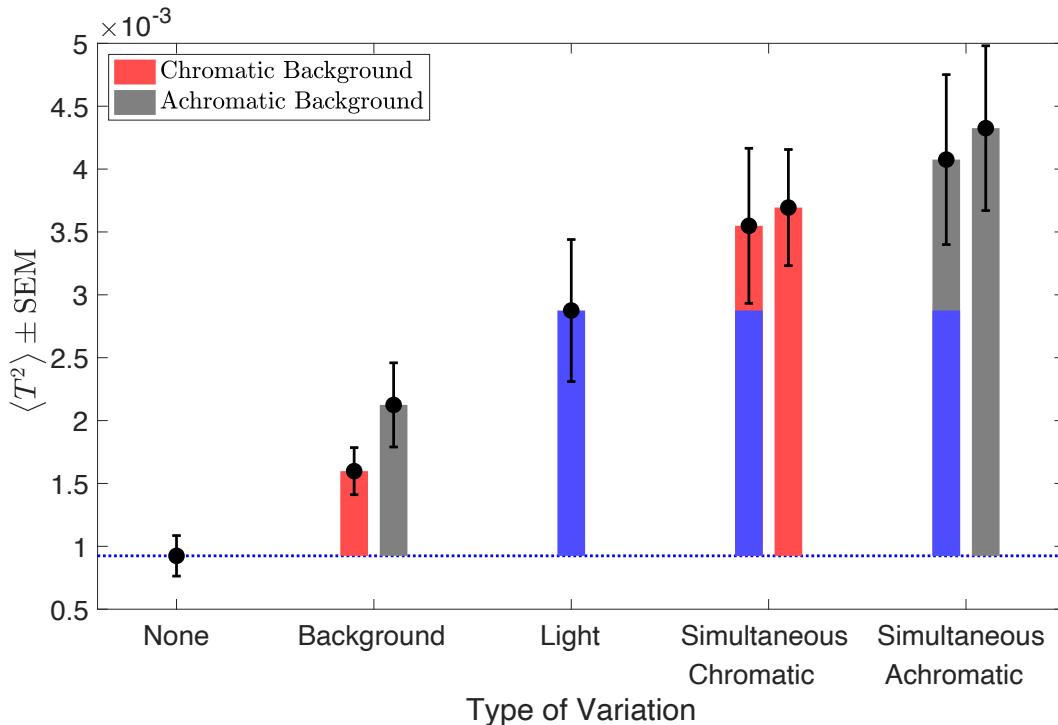
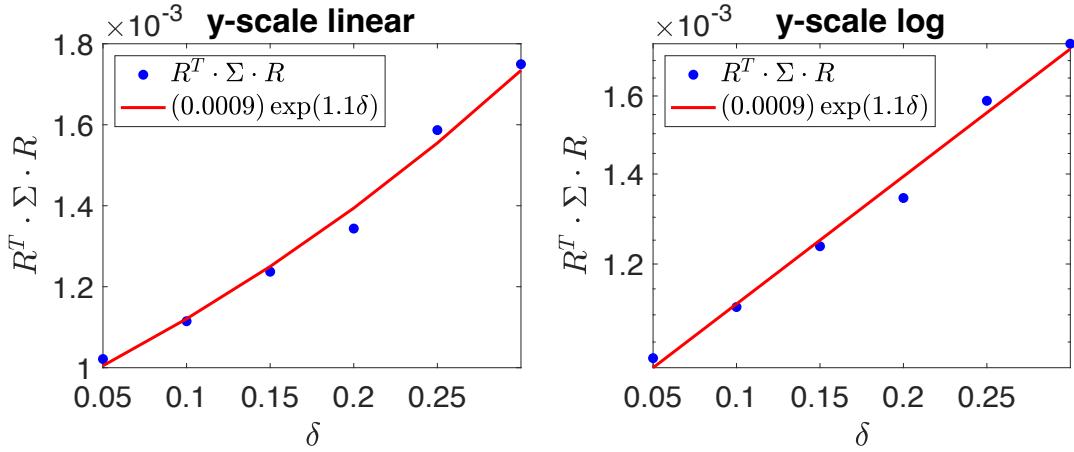


Figure 11: Discrimination thresholds for simultaneous variation of two sources are higher than individual discrimination thresholds. Mean ($N = 6$) log squared threshold for the six conditions in simultaneous variation experiment. The error bars represent ± 1 SEM taken between observers. The data for chromatic (red circles) and achromatic (gray diamonds) conditions have been plotted next to each other for visual comparison. The thresholds of the linear receptive field (LINRF) model (blue squares) were estimated using the parameters of the background variation condition (Figure 7) for the None, Background variation and Simultaneous variation conditions and using the parameters of the light intensity variation condition (Figure 9) for the Light condition. The blue error bars show ± 1 standard deviation estimated over 10 independent simulations. See Figure S7 for LINRF model thresholds with the same set of parameters for all conditions.

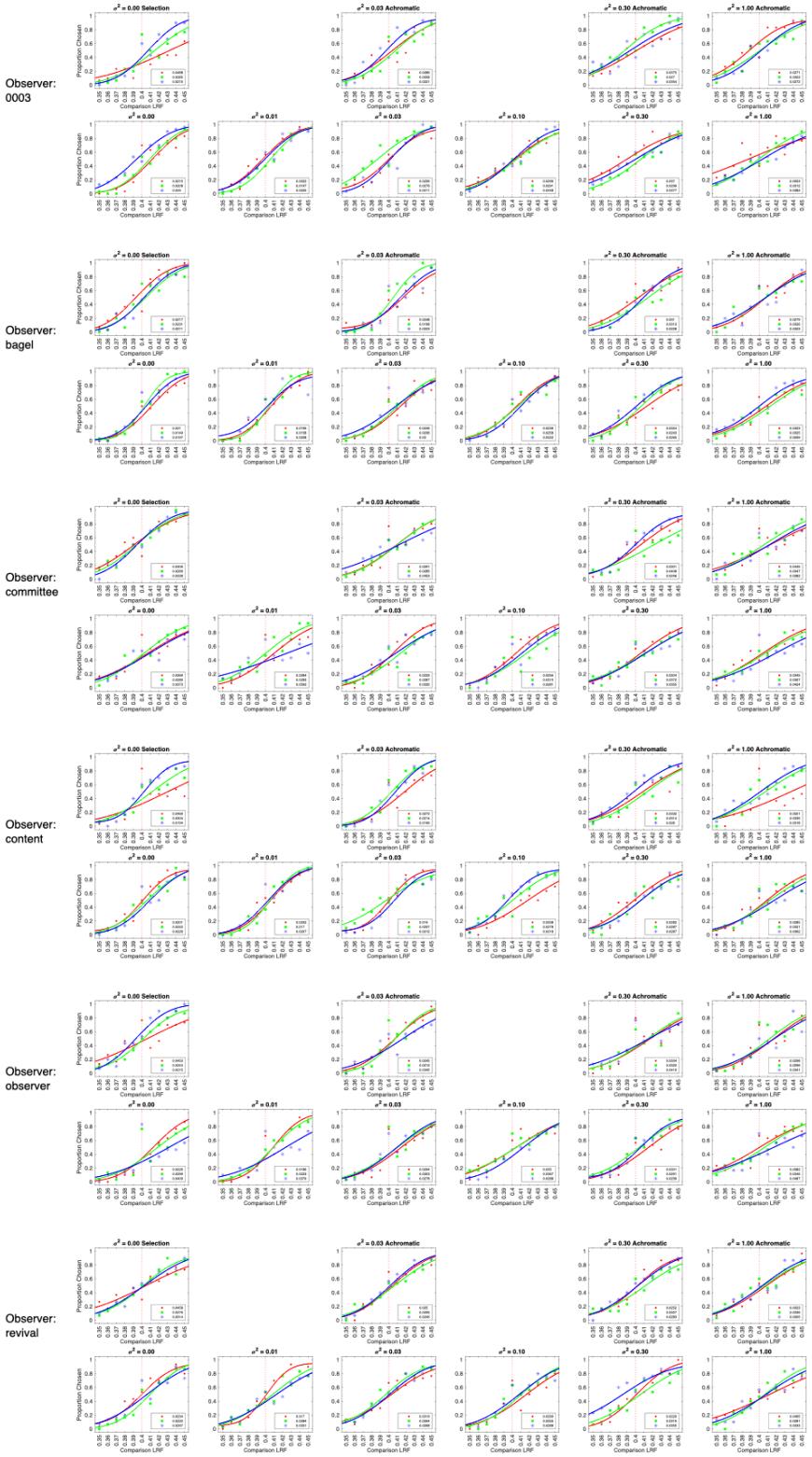


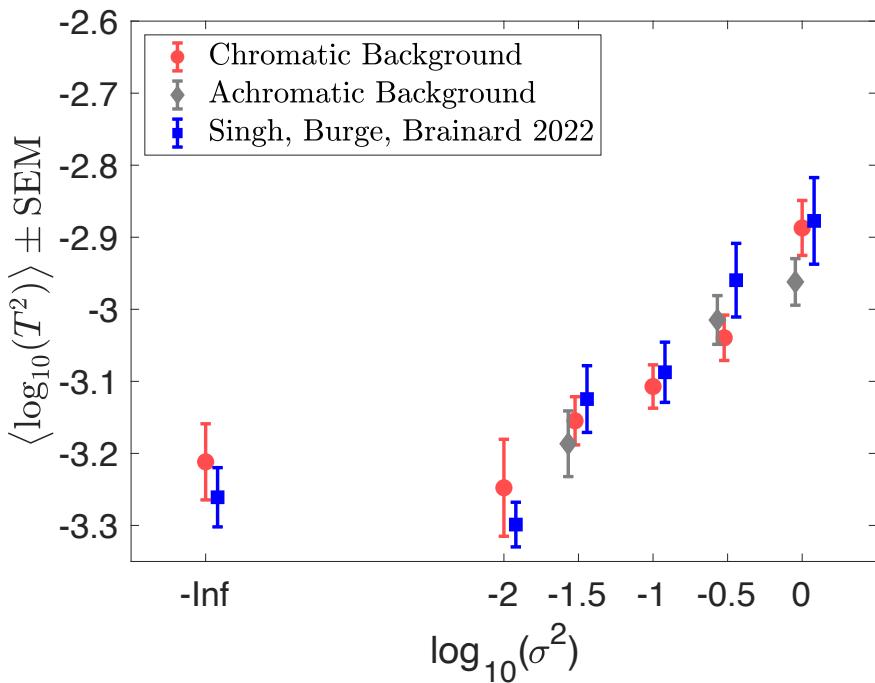
1

2 **Figure 12: Extrinsic noise of independent variations add linearly for simultaneous variation:** Mean
3 squared thresholds ($N=6$) for the six conditions in simultaneous variation experiment (black circles). The
4 black error bars represent ± 1 SEM taken between observers. The bars (red, gray, blue) represent the
5 increase in squared thresholds compared to the no variation condition (blue dotted line). For the
6 simultaneous variation conditions, the bars on the right (bars with one color, red or gray) represent the
7 increase in measured squared threshold for the simultaneous variation conditions and the bars on the left
8 (stacked bars of two different colors) represent the increase in the sum of the squared threshold of the
9 light intensity variation (blue bar) and the corresponding background variation conditions (red or gray).

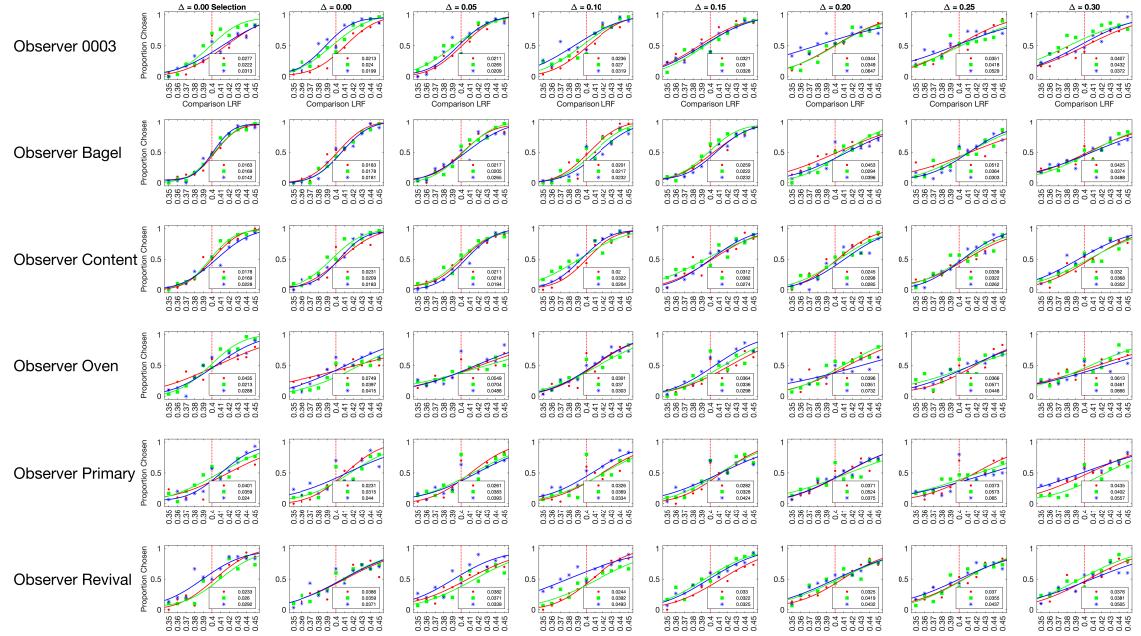


1
2 **Figure S1: Estimation of extrinsic noise for light intensity variation:** Plot of the variance ($R^T \Sigma R$) as a
3 function of the range parameter δ on a linear (left panel) and logarithmic (right panel) scale. We fit the
4 function with an exponential of the form $A * \exp(B * \delta)$. The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



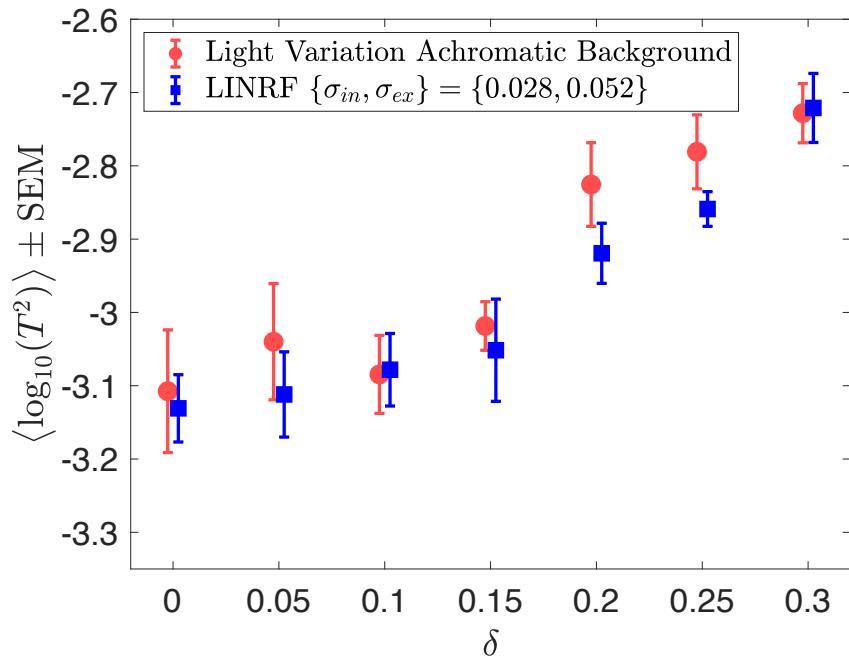


1
2 **Figure S3: Comparison with Singh, Burge, Brainard 2022.** Lightness discrimination thresholds for
3 background variation condition measured in preregistered Experiment 6 and previously reported data
4 from Singh, Burge, Brainard 2022. Preregistered Experiment 6 had both chromatic and achromatic
5 conditions, while the previous experiment (Singh, Burge, Brainard 2022) only had chromatic condition.
6 Singh, Burge, Brainard 2022 made three threshold measurements for each condition for 4 naïve
7 observers.



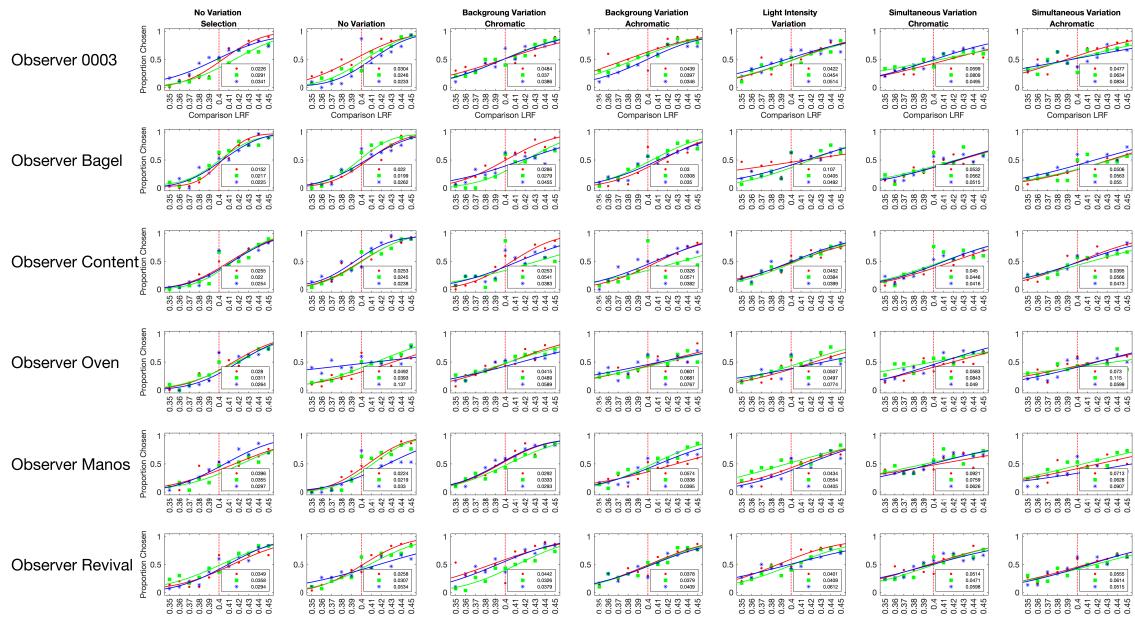
1

2 **Figure S4: Psychometric functions for all observers for light intensity variation experiment.** Same
3 as Figure 8, for all observers retained in light intensity variation experiment.



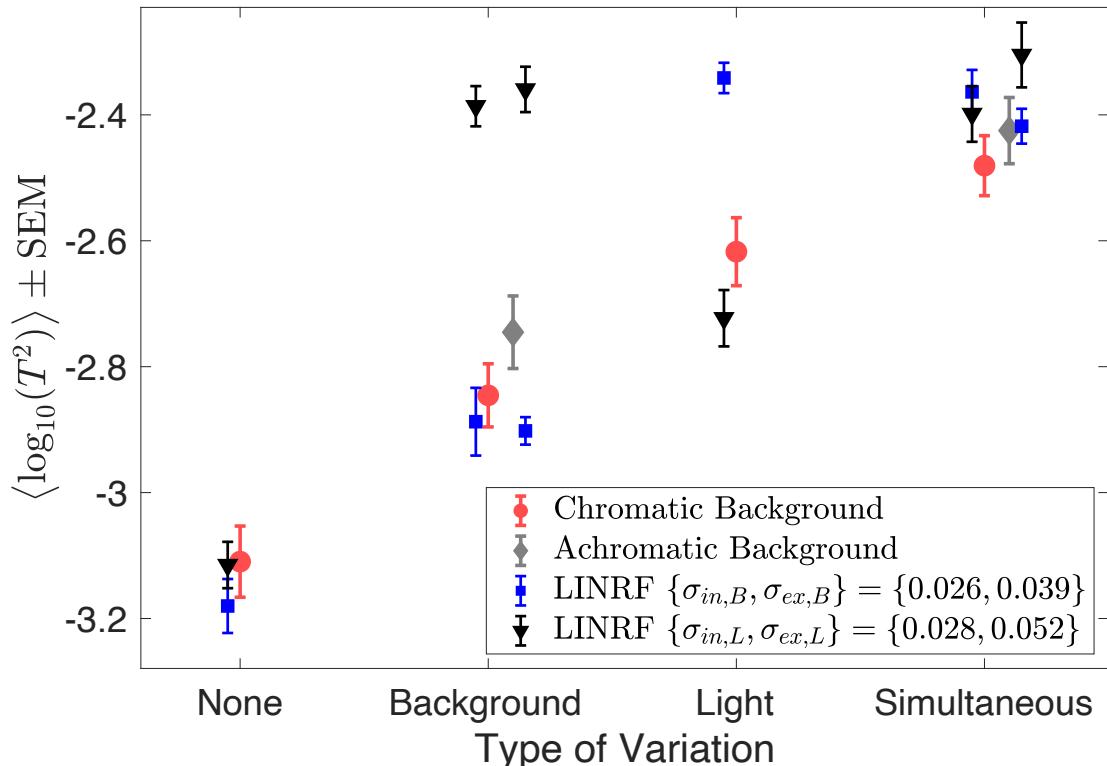
1

2 **Figure S5:** Same as Figure 9, for all six observers retained in light intensity variation experiment. The
3 parameters for the LINRF model are the same as in Figure 9.



1

2 **Figure S6: Psychometric functions for all observers for simultaneous variation experiment.** Same as
3 Figure 10, for all observers retained in simultaneous variation experiment.



1
 2 **Figure S7:** Same as Figure 11, but the thresholds of the linear receptive field (LINRF) model were
 3 estimated using the same set of parameters for all six conditions studied in Experiment 8. Blue square
 4 markers show log squared thresholds estimated using the parameters of the background variation
 5 condition (Experiment 6, Figure 7). Black triangular markers show log squared thresholds estimated using
 6 the parameters of the light intensity variation condition (Experiment 7, Figure 9). The black and blue error
 7 bars show +/- 1 standard deviation estimated over 10 independent simulations. The parameters of the
 8 background variation condition (Experiment 6, blue squares) predict the thresholds of the no variation
 9 condition, the background variation condition, and the simultaneous variation condition quite well, but
 10 fail to predict the threshold of the light source intensity variation condition. Similarly, the parameters of
 11 the light source intensity variation condition (Experiment 7, black triangles) predict the thresholds of the
 12 no variation condition, the light source intensity variation condition, and the simultaneous variation
 13 condition quite well, but fail to predict the threshold of the background variation condition. This could
 14 possibly be because observers in the three experiments were different. Future work would aim at studying
 15 these conditions using the same set of observers.