Diminishing Returns in Perceptual Color Space – Now in Color

Emily Stark[1] and Terece L. Turton[2] and Roxana Bujack[2]

[1]Safeguards Science and Technology, Los Alamos National Laboratory, USA
[21]Information Sciences, Los Alamos National Laboratory, USA

Abstract

Recent work has proven the non-Riemannian nature of perceptual color space by identifying the existence of diminishing returns along the luminance axis. A lurking question is if the luminance channel might somehow be the only axis that exhibits diminishing returns. In this short paper, we report on a companion study along two color axes: green to pink and blue to orange. These paths through color space were chosen as the most likely ones to form geodesics based on maximal agreement between color similarity experiments and hue constancy experiments. Our crowdsourced studies confirmed the existence of diminishing returns along both lines of color studied, bolstering the evidence that perceptual color space is non-Riemannian.

CCS Concepts
• ***Human-centered computing*** → ***Empirical studies in visualization; Visualization theory, concepts and paradigms;***

1. Introduction

A mathematical representation of the relationship between two colors has been a goal for centuries. Although a convenient mathematical construct, a Euclidean space is not a perceptual color space–one in which the distances between colors relate to how different humans perceive the colors to be. Over 100 years ago, Riemann [Rie21] and Schrodinger [Sch20] agreed that human perceptual color space was Riemannian. Riemannian geometry is a specific instance of differential geometry that includes real, smoothly varying manifolds. The shortest distance between two points in a Riemannian space is a geodesic. Judd [Jud79] introduced the concept of *diminishing returns* nearly 50 years ago. This refers to the phenomenon that the sum of small differences between two points adds up to more than the difference between the two points. Riemannian space is incompatible with the existence of diminishing returns.

Previous studies have established the existence of diminishing returns [Jud68, Jud70] along the luminance axis [BTM*22]. This indicates that there is at least one 1D subspace in perceptual color space which is non-Riemannian. Hence, perceptual color space must be non-Riemannian [Rie54, VH67, Sch20]. More recently, studies by Stark et al. [STM*24] validated the methodology by showing consistent diminishing returns along the gray axis between crowdsourced and in-person studies.

The next question is if diminishing returns exists elsewhere in color space and, if so, if it is constant throughout color space. Evidence has been collected towards consistency between the locations of geodesics in perceptual color space and its induced Riemannian space [BST*25]. Together with this result, constant diminishing returns throughout color space would imply that there exists a simple location-independent scaling function, $f : \mathbb{R} \to \mathbb{R}$ that maps an existing Riemannian distance to its diminished, non-Riemannian, perceptually accurate counterpart, such as a logarithm [HEL60] or a power law [Mac63], or a sine function [IS91]. This would also indicate that the underlying structure of perceptual color space is locally maintained in a Riemannian approximation.

We hypothesize that diminishing returns in perceptual color space is not constant, making a single, location-independent scaling function inaccurate in reproducing the perceived difference. Consequently, diminishing returns would need to be investigated and modeled across color space in order to capture perceptual uniformity. This would also mean that the underlying, non-Riemannian geometry is not so easily reconstructed from a Riemannian space. To address whether the effect of diminishing returns is consistent across color space, we repeat the triad stimulus methodology used in Bujack et al. [BTM*22]. Triads are constructed along two different color paths in CIELAB. To demonstrate diminishing returns, the middle color must lie on the shortest path between the two tests. Special care is taken to develop the triads so that the colors most likely lie on a geodesic. To this end, we use the theory that colors of constant luminance and hue form shortest paths [Sch20].

Based on work by Sanders and Wyszecki [SW57, WS57], isoluminance surfaces align well with the L^* axis in CIELAB; thus we choose constant L^* lines. Ebner and Fairchild [EF98] determined surfaces of constant hue in color space. These can be coupled with experiments underlying the 1976 and 2000 ΔE distance metrics. Because ΔE_{1976} is Euclidean, the shortest paths in this metric are always straight lines. To find paths that are most likely geodesics,

DOI: 10.2312/evs.20251081

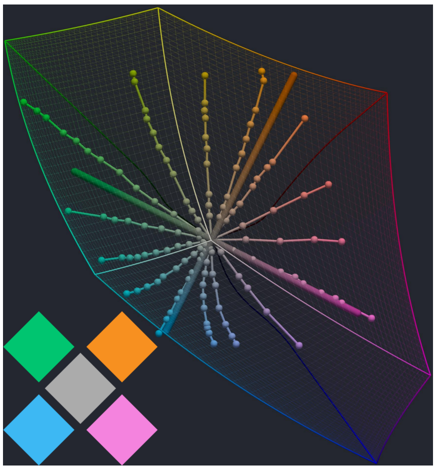

Figure 1: *Results from hue-constancy experiments [EF98] linearly interpolated and intersected with $L^* = 70$ in thin lines with points embedded in CIELAB plotted with the sRGB gamut as boundaries. Thick lines show the lines used for our experiments. Inset: Endpoints of the two straight and approximately perpendicular lines.*

we look for regions where the geodesics in ΔE_{2000} and the iso-hue lines are as straight as possible, so that ΔE_{2000}, ΔE_{1976} and the Ebner Fairchild data are all in agreement. Each distance metric forms the same paths in the a^*b^* plane independent of L^*. Thus iso-hue surfaces [EF98] are the determining factor for the selection of L^*. For all pairs of orthogonal lines, the best agreement across all three measures, ΔE^*_{76}, ΔE^*_{00}, and the [EF98] data, are at $L^* = 70$, $a^* = -2b^*$ and $b^* = 2a^*$. Fig. 1 shows the ΔE^*_{00} geodesics, computed using 2D relaxation [BCT*24], on top of the iso-hue lines. Along each of these selected lines, three centers are chosen to maximize the space between them. Line 1 is defined as: green (HEX: #00C570, $L^* = 70, a^* = -60, b^* = 30$) to pink (HEX: #F483DE, $L^* = 70, a^* = 61.33, b^* = -30.67$). Line 2 is defined as: blue (HEX: #3DB8F3, $L^* = 70, a^* = -20, b^* = -40$) to orange (HEX: #F79020, $L^* = 70, a^* = 34.74, b^* = 69.49$), Figure 1.

It is predicted that the effect of diminishing returns along these two lines will not be identical. To evaluate this, maximum likelihood estimation (MLE) [MY03, KM08, TTMB22] approximation of the scaling functions will be compared, using the same value for standard deviation. If diminishing returns is inconsistent across color space, a model of diminishing returns approximated using triads along one line will fail at predicting responses to triads along the other. For additional reading on the history and the geometry of color space, we recommend the works by Resnikoff [Res74], Vos [Vos79], and Wyszecki and Stiles [WS82].

2. Method

Triads are constructed by sampling colors along the selected lines. Approximately equal steps in ΔE^*_{00} were used to generate the color sets. These colors were used to design experimental triads, selecting three equally spaced centers along each line.

As in Bujack et al. [BTM*22], this study was completed via Qualtrics with participants recruited on MTurk. Crowdsourcing has been used as a research platform for well over a decade in visualization [BMB*18] and perception research [ACC*15]. In a recent publication, Stark et al. [STM*24] directly compared an in-person study with the Bujack et al. [BTM*22] crowdsourced study, validating the crowdsourced methodology and demonstrating that the variations in experimental setup can be averaged out as noise. Participants were treated in accordance with the Los Alamos National Laboratory Human Subject Research Review Board. Each gave consent, self-identified as not colorblind, passed an Ishihara plate text, read instructions and did three training tasks. Participants were restricted to those that had completed the Bujack et al. [BTM*22] study with significantly above chance accuracy. Participants were shown a triad with the standard in the middle and the tests on the right and left, horizontally. Participants indicated which test was *more different* by responding with the "q" or "p" key to select the left or right test. Participants were told whether they selected the correct test after each trial.

A total of 755 participants completed this study. Each was shown a random subset of 60 triads. Each experimental triad had approximately 120 responses ($M = 119.8, SD = 10.1$). Participants' ages ranged from 18 to 80, with a mean of 39.5. 53.5% identified as male, 45.4% identified as female, and 1.3% as other or no answer. Participants were compensated at a level consistent with the median participant time paid at least the federal hourly minimum wage. Additional study details are in the supplemental material.

3. Results

Inference Testing

Before performing the MLE, the existence of diminishing returns was confirmed using the regression model to predict participant accuracy based on the difference of differences Δd and the average difference per triad, $\bar{d}$. Using the same framework as in Bujack et al. [BTM*22], except now Δd is treated as a continuous variable, the regression model is

$$\widehat{acc} = \beta_0 + \beta_{\Delta d} * \Delta d + \beta_{\bar{d}} * \bar{d} + \beta_{\Delta d * \bar{d}} * \Delta d * \bar{d}. \quad (1)$$

Separate regressions were performed for the two lines. Coefficient values can be seen in Tables 1 and 2. These models were fit using scaled values for the predictors. These results indicate that as the average difference increases, accuracy drops along both lines, $\beta_{\bar{d},line1} = -0.376, t(113) = -5.108, p < 0.001$, $\beta_{\bar{d},line2} = -0.483, t(119) = -7.102, p < 0.001$. These models also achieved an acceptable goodness-of-fit. Both the model for line 1, $R^2_{adj} = 0.5119, F(3,185) = 66.74, p < 0.001$, *Cohen's* $f^2 = 1.049$, and line 2, $R^2_{adj} = 0.5373, F(3,185) = 38, p < 0.001$, *Cohen's* $f^2 = 1.161$, achieved a large effect size. Hence, there is evidence for the existence of diminishing returns. The relationship between average difference and accuracy can be seen in Figure 2, where there is a negative trend.

MLE Modeling for the Existence of Diminishing Returns

Four models were learned using MLE and their ability to predict participants' responses were compared. The first is a baseline condition where ΔE^*_{00} was scaled linearly,

$$\Delta d_t = \frac{\Delta E^*_{00}(stand, test1) - \Delta E^*_{00}(stand, test2)}{c}. \quad (2)$$

Table 1: *Coefficients for green-pink regression model*

	Line 1 (Green-Pink)	
Term	Coefficient	P-value
Intercept	-0.003	0.948
Effect of Δd	0.635	< 0.001
Effect of $\bar{d}$	-0.277	< 0.001
Interaction of Δd and $\bar{d}$	-0.063	0.371

Table 2: *Coefficients for orange-blue regression model*

	Line 2 (Orange-Blue)	
Term	Coefficient	P-value
Intercept	0.002	0.957
Effect of Δd	0.604	< 0.001
Effect of $\bar{d}$	-0.349	< 0.001
Interaction of Δd and $\bar{d}$	0.102	0.139

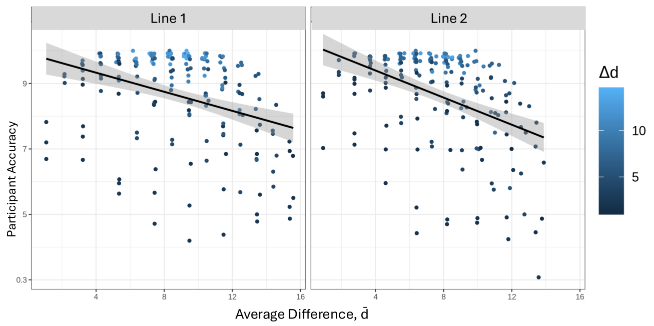

Figure 2: *The relationship between average difference and accuracy. Color corresponds to the difference of difference. All differences are calculated using* ΔE^*_{00}.

The second model applies the difference scaling function, $f(x)$, which takes in ΔE^*_{00},

$$\Delta d_t = f\left(\Delta E^*_{00}(stand, test1)\right) - f\left(\Delta E^*_{00}(stand, test2)\right). \quad (3)$$

The third model is a perceptual function, $g(x)$, that transforms the raw colors into the perceived strength, Ψ, applied to the a^* value,

$$\Delta d_t = |g(a^*_{stand}) - g(a^*_{test1})| - |g(a^*_{stand}) - g(a^*_{test2})|. \quad (4)$$

The final model applies a difference scaling function to the data after applying the perceptual function,

$$\Delta d_t = f\left(|g(a^*_{stand}) - g(a^*_{test1})|\right) - f\left(|g(a^*_{stand}) - g(a^*_{test2})|\right). \quad (5)$$

Cross-validation and bootstrapping were used when training the models. Ten test/train splits were used. For each split, a random selection of 20 responses per triad was held out as the test set. The remaining responses were aggregated in the full training set, from which 100 bootstrapped training sets of size 18900 were sampled. For each bootstrap step, the learned models were evaluated on the test set to determine how well they predict participants' responses. Models were also learned using data from only one line.

The average accuracy of each model applied to the test data can be seen in Figure 3. Along line 1 (green-pink), there is an increase in accuracy when applying the $g(x)$, compared to the baseline condition. This indicates that ΔE^*_{00} is insufficient to describe the perception of color differences, assuming additivity. After applying the perceptual scale, there is another significant increase in accuracy when allowing for diminishing returns. This suggests that even after correcting for the perceptual scale, there is evidence for the nonadditivity of small differences.

Along line 2 (orange-blue), the pattern of responses is different. Applying a perceptual scale does not significantly increase predictive power, compared to the baseline. However, when accounting for nonadditivity of small differences, there is an increase in accuracy both when comparing the baseline to the $f(x)$ only models and comparing $g(x)$ only to $g(x)$ and $f(x)$. This supports the existence of diminishing returns along this region in CIELAB as well.

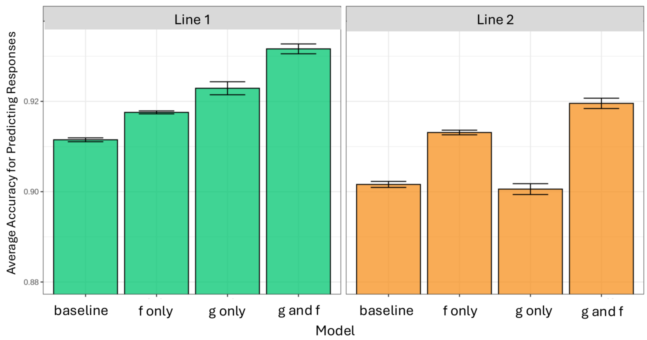

Figure 3: *Predictive power of the four models. All models were trained using data and predicted responses along the same line.*

The perceptual functions are plotted against the a^* values for line 1 and b^* values for line 2 in Figure 4 where the shaded region indicates the interquartile range for learned models. Due to the linear dependence between a^* and b^* along both lines, the choice is arbitrary as it only changes the $x-$values of the control points. These axes were chosen as they both have the largest domains so the perceptual functions can be more easily compared. The most notable difference is the symmetry that is present for the perceptual scale of line 1, which does not exist in the model for line 2. This may contribute to the decreased accuracy along line 2 when applying the perceptual function. The asymmetry of the model along line 2 is likely due to the asymmetric nature of the triads. Due to gamut limitations, triads were centered at $a^* = -5.7, b^* = -11.3$; $a^* = 1.9, b^* = 3.9$; and $a^* = 13, b^* = 26.1$. These centers were selected as they were approximately equidistant along the perceptible region of line 2. To further investigate this, we compared the accuracy of the four models based on their location in CIELAB. The baseline model is more accurate for both models when a^* is near 0, while the models that accommodate the perceptual scales are less accurate there. This suggests that ΔE^*_{00} is better suited for less saturated colors. It also demonstrates a tradeoff in this analysis to accurately model perceived differences of colors compared to a near-gray standard versus a more saturated standard. This may be mitigated by increasing the number of control points in the spline, but that would likely require much more data based on our results from the validation of these methods. The scaling functions learned

Proceedings published by Eurographics - The European Association for Computer Graphics.

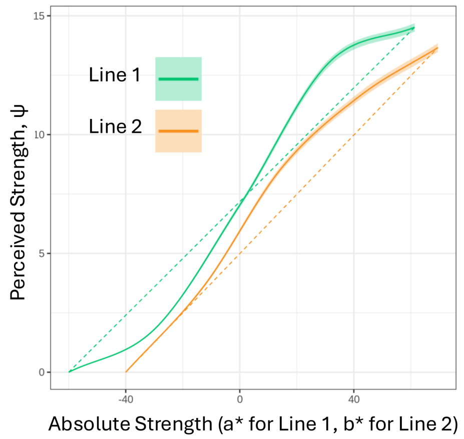

Figure 4: *Learned perceptual scales g of both lines.*

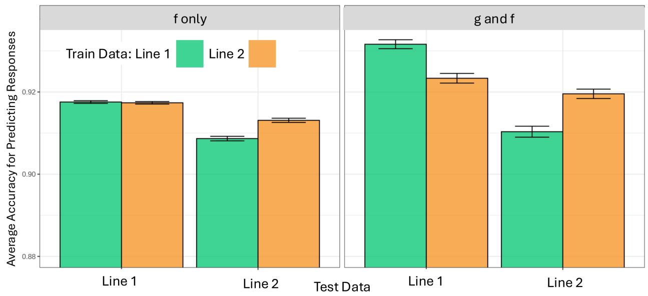

Figure 6: *Test accuracy based on which path through color space was used to learn the MLE (train data) and which path was used to test the model (test data).*

with and without a perceptual function are seen in Figure 5. Regardless of the application of a perceptual function, the difference scaling functions are concave. The dashed lines indicate what would have been learned if there were no diminishing returns.

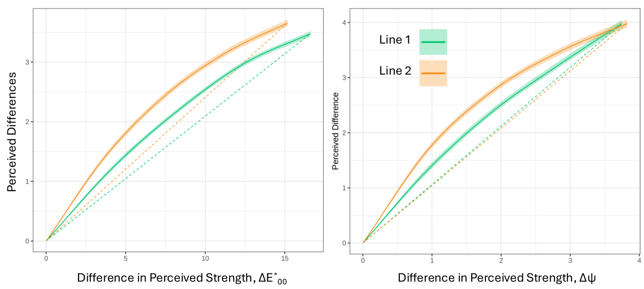

Figure 5: *Learned difference scaling functions f of both lines.*

Comparing Diminishing Returns Between Regions

Diminishing returns exists along both paths in CIELAB. The question remains whether diminishing returns exists to the same extent. The learned difference scaling functions, $f(x)$, can be directly compared because the inputs are on the same scale, consistent with $\sigma = 1$. The scaling function learned from data along one line is compared to the test set from the other line. This was done for the case of $f(x)$ only and for both $g(x)$ and $f(x)$. The results, seen in Figure 6, indicate that accounting for the diminishing returns as it exists along line 1 does not yield the same predictive power when applied to line 2. If diminishing returns existed to the same extent along the two lines, there would not be a significant difference between the accuracy within testing data. Combined with the lack of overlap between the models seen in Figure 5, we conclude that diminishing returns is not consistent through color space.

4. Discussion

The results from Bujack et al. [BTM*22] and this work strongly support diminishing returns in the luminance channel and along two color paths in CIELAB. The existence of this phenomenon implies that small differences are nonadditive, which cannot be accommodated in a Riemannian space. Using the scaling function learned from one line applied to the other, there was a decrease in accuracy, indicating that diminishing returns are not constant across color space. The decrease in accuracy when modeling the perceived strengths along line 2 is unexpected. However, the model with the highest accuracy along line 2 was that which used both a perceptual function and a difference scaling function. These functions were estimated separately so the decreased accuracy when applying only $g(x)$ was not a result of fitting the $f(x)$.

Additionally, accounting for the size of the differences with a concave function significantly improved the ability to predict participants' responses, consistent with the existence of diminishing returns. Because the difference scaling functions varied based on the region in CIELAB, diminishing returns are not consistent across all color space. This finding contradicts the hypothesis that the perceptual metric is just a scaled version of a Riemannian metric, with a monotonic scaling function only dependent on the distance between the points. These results indicate that both the size of the difference *and the location* are required to accurately quantify the perceptual metric, or how different colors are perceived to be.

The finding that diminishing returns exists at different strengths along two paths in color space is groundbreaking. Previous work implicitly assumes that color space can be modeled simply through a 1D scaling function [Mac63, HEL60, IS91, BTM*22]. It is, for example, possible that knowledge about the LMS [Sto19] differences of two points is sufficient to infer the effect size of diminishing returns, i.e., that the metric takes a shape of the form $f(f_L(\Delta L) + f_M(\Delta M) + f_S(\Delta S))$ with four 1D scaling functions. However, this hypothesis cannot be tested on measurements from two geodesics only. So far, our findings suggest that a truly perceptually uniform color space requires not only a non-Riemannian geometry but an extensive investigation of how diminishing returns changes throughout color space.

Acknowledgements Research presented in this article was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project number 20200512ECR and released under LA-UR-25-21303.

References

[ACC*15] AT W., C V., CA L., X W., C. S.: Conducting perception research over the internet: a tutorial review. *PerrJ* (2015). doi:10.7717/peerj.1058. 2

[BCT*24] BUJACK R., CAFFREY E., TETI E., TURTON T. L., ROGERS D. H., MILLER J.: Efficient computation of geodesics in color space. *IEEE Transactions on Visualization and Computer Graphics 30*, 9 (2024), 6507–6519. doi:10.1109/TVCG.2023.3346673. 2

[BMB*18] BORGO R., MICALLEF L., BACH B., MCGEE F., LEE B.: Information Visualization Evaluation Using Crowdsourcing. *Computer Graphics Forum* (2018). doi:10.1111/cgf.13444. 2

[BST*25] BUJACK R., STARK E., TURTON T. L., MILLER J., ROGERS D. H.: The Geometry of Color in the Light of a Non-Riemannian Space, 2025. To appear in proceedings of the Eurographics Association. 1

[BTM*22] BUJACK R., TETI E., MILLER J., CAFFREY E., TURTON T. L.: The non-riemannian nature of perceptual color space. *Proceedings of the National Academy of Sciences 119*, 18 (2022). doi:10.1073/pnas.2119753119. 1, 2, 4

[EF98] EBNER F., FAIRCHILD M. D.: Finding constant hue surfaces in color space. In *Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III* (1998), vol. 3300, International Society for Optics and Photonics, pp. 107–117. doi:10.1117/12.298269. 1, 2

[HEL60] HELM C.: *A successive intervals analysis of color differences*. Tech. rep., PRINCETON UNIV NJ, 1960. 1, 4

[IS91] IZMAILOV C. A., SOKOLOV E.: Spherical model of color and brightness discrimination. *Psychological science 2*, 4 (1991), 249–260. doi:10.1111/j.1467-9280.1991.tb00143.x. 1, 4

[Jud68] JUDD D. B.: Ideal color space: Curvature of color space and its implications for industrial color tolerances. *Palette 29*, 21-28 (1968), 4–25. 1

[Jud70] JUDD D. B.: Ideal color space. *Color Engineering 8*, 2 (1970), 37. 1

[Jud79] JUDD D. B.: *Contributions to color science*, vol. 545. Department of Commerce, National Bureau of Standards, 1979. 1

[KM08] KNOBLAUCH K., MALONEY L. T.: Mlds: Maximum likelihood difference scaling in r. *Journal of Statistical Software 25*, 2 (2008), 1–26. doi:10.18637/jss.v025.i02. 2

[Mac63] MACADAM D. L.: Nonlinear relations of psychometric scale values to chromaticity differences. *JOSA 53*, 6 (1963), 754–757. 1, 4

[MY03] MALONEY L. T., YANG J. N.: Maximum likelihood difference scaling. *Journal of Vision 3*, 8 (2003), 5–5. doi:10.1167/3.8.5. 2

[Res74] RESNIKOFF H. L.: Differential geometry and color perception. *Journal of Mathematical Biology 1*, 2 (1974), 97–131. doi:10.1007/BF00275798. 2

[Rie54] RIEMANN B.: Uber die hypothesen, welche der geometrie zu grunde liegen. *Königliche Gesellschaft der Wissenschaften und der Georg-Augustus-Universität Göttingen 13*, 133 (1854), 1867. 1

[Rie21] RIEMANN B.: Über die hypothesen, welche der geometrie zugrunde liegen. In *Über die Hypothesen, welche der Geometrie zu Grunde liegen*. Springer, 1921, pp. 1–47. 1

[Sch20] SCHRÖDINGER E.: Grundlinien einer theorie der farbenmetrik im tagessehen. *Annalen der Physik 368*, 22 (1920), 481–520. 1

[STM*24] STARK E. N., TURTON T. L., MILLER J., BARENHOLTZ E., HONG S., BUJACK R.: Toward the validation of crowdsourced experiments for lightness perception. *PLoS ONE 19*, 12 (2024). doi:10.1371/journal.pone.0315853. 1, 2

[Sto19] STOCKMAN A.: Cone fundamentals and cie standards. *Current Opinion in Behavioral Sciences 30* (2019), 87–93. doi:10.1016/j.cobeha.2019.06.005. 4

[SW57] SANDERS C., WYSZECKI G.: Correlate for lightness in terms of cie-tristimulus values. part i. *JOSA 47*, 5 (1957), 398–404. 1

[TTMB22] TETI E. N., TURTON T. L., MILLER J. M., BUJACK R.: Maximum likelihood estimation of difference scaling functions for suprathreshold judgments. *Journal of Vision 22*, 10 (2022), 9–9. doi:10.1167/jov.22.10.9. 2

[VH67] VON HELMHOLTZ H.: *Handbuch der physiologischen Optik*, vol. 9. Voss, 1867. 1

[Vos79] VOS J.: Line elements and physiological models of color vision. *Color Research & Application 4*, 4 (1979), 208–216. 2

[WS57] WYSZECKI G., SANDERS C.: Correlate for lightness in terms of cie-tristimulus values. part ii. *JOSA 47*, 9 (1957), 840–842. 1

[WS82] WYSZECKI G., STILES W. S.: *Color science*, vol. 8. Wiley New York, 1982. 2