

THE UNDERLYING DISTRIBUTION OF LIGHTNESS MATCHES

Suncica Zdravkovic and Petar Milin

*Department of Psychology, University of Novi Sad
Laboratory for Experimental Psychology, University of Belgrade
szdravko@f.bg.ac.yu, pmilin@ff.ns.ac.yu*

Abstract

The normality assumption for lightness matches was tested. A large number of observers were run on the well-known simultaneous lightness contrast illusion and the distribution of the obtained data was analysed. The lightness matches were gathered using the Munsell scale. The distribution testing was done both for the reflectance data and data transformed into log reflectance. The tests showed supporting evidence in the case of reflectance data. They could be treated as normally and double-exponentially (Laplace) distributed. However no such symmetrical distribution can be assumed for the log transformed reflectance values. Results strongly suggest that the use of standard statistical methods should be limited only to reflectance values. Some alternative methods are suitable in case of Laplace distribution and can be used as well.

The assumption that phenomena are normally distributed in the population is called the normal distribution assumption. It is ubiquitous in psychological studies including those relating to attitudes, preferences, personality traits, intelligence and so forth. The normal distribution assumption is crucial because it allows the usage of various statistical techniques in data analysis.

In perception however, phenomena are not necessarily expected to be normally distributed. The requirement for normality of data distribution is therefore fulfilled in a different fashion. Each observer in a perception experiment typically repeats the same task a number of times and the mean value of all repetitions is used. This procedure is quite different from the procedures mentioned above (i.e. a subject does not repeatedly respond to the same questionnaire- or test-item). The rationale behind the repetition in perception experiments is the following: no matter what the initial distribution, the mean values will always tend to be normally distributed (justifying the central limit theorem). This is easily illustrated with a dice. When a single dice is thrown the probability for it to fall on any side is $1/6$. Therefore the distribution resulting from such a demonstration will be a line parallel with x (and $y = n/6$). If two dices are thrown and the mean value is taken, the distribution will become normal. Similarly, the mean value of several responses in a perception experiment will produce a normal distribution.

In the field of lightness perception we encounter a specific problem concerning this methodology. In perceptual experiments run on computers it is easy to change the stimulus and to present the same display more than once. In lightness perception, real objects under real illumination are used (from Katz, 1911 to Robilotto & Zaidi, 2006). In such conditions it is not easy to switch from one stimulus to another without the observer noticing it. In fact, most of the time the experimental set up does not permit running the same subject on the same stimulus more than once. Even when switching is possible the objects tend to keep their identity in a manner very different from CRT presented stimuli. Therefore it is difficult to

reintroduce the same object without the observer being aware of the repetition. Once an observer recognizes the previously used stimulus she is no longer naïve for the purposes of the immediate task repetition. Therefore in lightness perception we typically have a single match from each observer.

The present study deals with the question of the underlying distribution of the data obtained when only a single measure is taken. Are those data normally distributed, although the mean values are not used? Or are the data distributed in some other manner and if so, how? Will such a distribution still be symmetrical, with known or estimated parameters? From a statistical point of view it is crucial to be aware of the underlying distribution as it dictates what statistical techniques can be used.

To answer large issues, such as the problem of underlying distributions, meta-analysis is often a useful approach. However, as determining distributions is sensitive to the slight shifts in values, summing the results for several different studies was not an option in our case. The required precision must come from measurements taken under the very same experimental conditions. Therefore we tested 61 different observers, which is an unusually large number for this type of perceptual study.

In this study we used the well-known phenomenon, illusion of lightness simultaneous contrast. The display consists of two identical middle-grey targets placed on the different backgrounds, black and white (figure 1a). The target against the black background appears to be much lighter. This illusion holds over a number of conditions (figure 1b-e): when the targets are on paper or on CRT (Gilchrist et al. 1999), when the whole illusion display is intensely illuminated (Agostini & Bruno 1996), when the backgrounds are closer in shade than just black and white (Arend & Goldstein, 1987; Bressan & Actis-Grosso, 2001), when the backgrounds are no longer uniformly grey (Schirillo & Shevell, 1996; Bruno, Bernardis, & Schirillo, 1997), with gradient backgrounds (Bressan, 2003), with articulated backgrounds (Cannon & Fullenkamp, 1991; Adelson, 2000; Salmela & Laurinen 2005; Bressan & Actis-Grosso, 2006) or with static and dynamic noise (Zdravkovic et al, 2005), or when the targets have added frames (Gilchrist, 2006).

Method

Subjects: 61 first-year students from Belgrade psychology department participated in this study for partial class credit. All of them had normal or corrected to normal vision. As the after experiment debrief showed, the observers were already familiar with the illusion. However, the theoretical familiarity with the phenomenon does not diminish the perceptual effect (which was reconfirmed in this study).

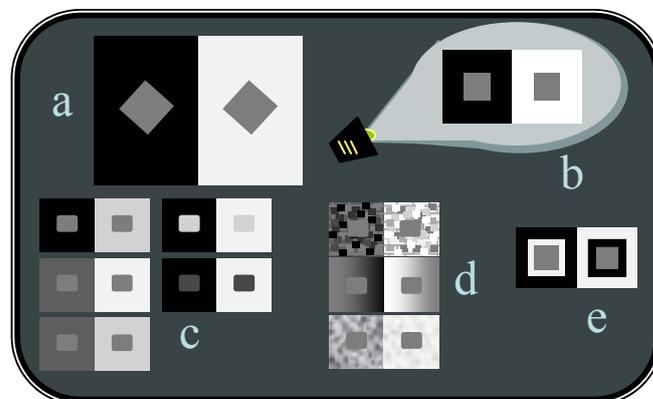


Figure 1. Versions of simultaneous lightness contrast: a) original display, b) under intense illumination, c) variation of grey backgrounds and targets, d) articulation, e) frames added.

Table 1. The reflectance and luminance values for the three surfaces in the experiment

	Target	Light background	Dark background
Colour-Aid paper	4.5	White	Black
Reflectance (%)	15.6	90	3.1
Luminance (cd/m ²)	1.2	7	0.24

Stimulus: The display (figure 1a) was made of Colour-Aid paper, a specially manufactured Lambertian surface with the certified reflectance value. Three different shades were used to create the display (table 1). The targets were 7cm x 7cm each and the backgrounds were 20cm x 29 cm.

The display was attached to the inside wall of the chamber (75cm x 75cm x 150cm), which ensured both the control of illumination and of viewing conditions (figure 2). The chamber was illuminated with a single light source (light bulb, 220V, 60W) placed above observer's head.

The light reflected from each surface is called luminance and in physics it is the product of reflectance and illumination. The luminance values for all surfaces are given in Table 1.

Observer looked inside the chamber while making the perceptual judgements. The observer was 150 cm away from the display, with her head at a fixed position. From this viewpoint the target subtended 2.67° and the background 7.64°x11.08° of visual angle.

Observers were asked to make a lightness judgement that is to judge the reflectance of the two targets. The order of surfaces that were judge was randomised across the observers. The viewing time was not limited and it took observer about two minutes to finish the task. The observer task was to match the surface on the display with one of the grey shades on the Munsell chart. This chart is standardized tool for lightness matching: it is a 16-step scale of grey shades varying in reflectance from 3% (black) to 90% (white).

Statistical procedure: Beside standard ANOVA that were used for confirming well-known general findings, testing for the distribution assumptions was three-folded.

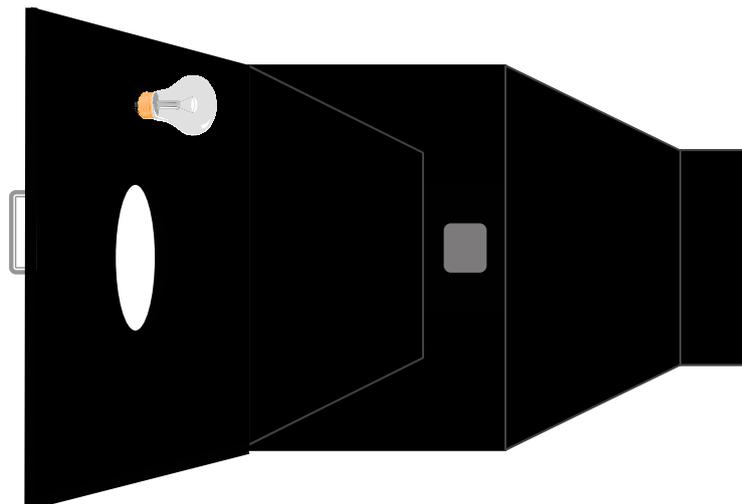


Figure 2. The chamber with the display on the far wall and the opening for the observer's head on the near wall. The light source is placed above the observer's head and the chamber is closed during the experiment.

First, a set of one-sample tests were applied, including Cramer-von Mises (W^2), Watson (U^2), Anderson-Darling (A^2), Kolmogorov-Smirnov (D) and Kuiper (V). The power studies of Puig and Stephens (2000) have shown that W^2 , U^2 and A^2 are generally more powerful. The Watson (U^2) statistic is particularly useful for the problem of testing symmetric distributions, like normal, lognormal and Laplace. However, D and V are standard in various packages for statistical analysis and, hence, better known.

Second, Bayesian inference (Sivia, 1996; Winkler, 2003; MacKay, 2003), where log-likelihood ratios of the two competing distributional hypothesis (normal vs. lognormal vs. Laplace) has been calculated as the evidence in favour of one or the other distribution. Without going into technical details, let $P(\text{Hy-normal})$, $P(\text{Hy-lognormal})$ and $P(\text{Hy-Laplace})$ be the *prior probabilities* (i.e. prior evidence) for each of the hypothesized distributions (usually, two are equally probable: 0.5 – 0.5). Given a collected set of points $X = \{x_1, \dots, x_N\}$, m and s would be the estimates of mean and standard deviation, obtained by integration over the parameter space. Then:

$$L1(m, s) = \log [P(\text{Hy-normal} | X, m, s) / P(\text{Hy-Laplace} | X, m, s)] \quad (1)$$

$$L1(m, s) = \log [P(\text{Hy-normal} | X, m, s) / P(\text{Hy-lognormal} | X, m, s)] \quad (2)$$

$$L2(m, s) = \log [P(\text{Hy-lognormal} | X, m, s) / P(\text{Hy-Laplace} | X, m, s)] \quad (3)$$

would provide a direct, analytical way to conclude how much more likely (in log units) one distribution is than the other given the data. For example, if $L1 > 0$ then there is more evidence in favour of normal distribution, while if $L1 < 0$ Laplace is favourite, and so on.

Third, bootstrapping of the Kolmogorov-Smirnov one-sample test has been done (Abadie, 2002; Diamond & Jasjeet, 2005). In 1000 replications, one-sample D-statistics was calculated between observed (empirical) data and randomly generated data, distributed, normally, lognormally or double-exponentially (Laplace), with observed position and scale parameters (appropriately, mean or median; standard deviation or absolute deviation). Results include the bootstrap p-values of the Kolmogorov-Smirnov test for the hypothesis that the probability densities for both empirical (observed) and theoretical (expected normal, lognormal or Laplace) are the same.

RESULTS

Our results reconfirmed the standard finding both for the direction and the strength of the illusion. The target on the dark background appeared lighter with 1.3 Munsell steps difference between the two targets (Figure 3a).

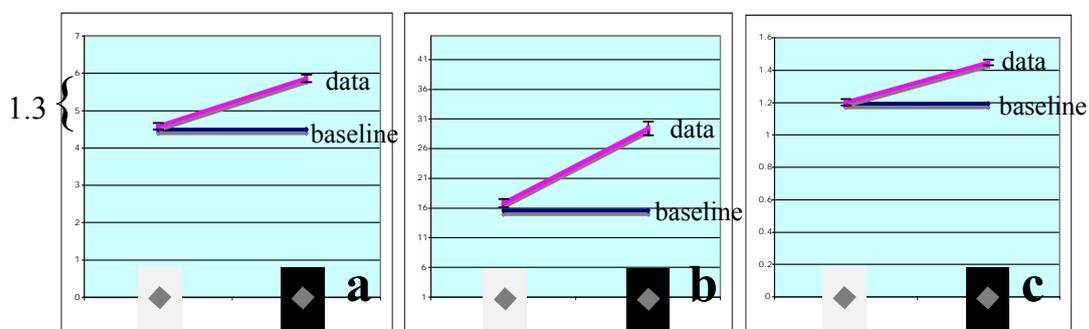


Figure 3. The obtained results presented with a) Munsell values on y; b) reflectance on y; c) log reflectance on y. Although the corresponding range of values is presented on y, the results appear different.

Table 2. The distribution analysis for obtained reflectance values and critical values.

Background	Distribution	Test				
		W ²	U ²	A ²	D	V
Target on white	normal	0.271	0.265	1.513	1.397	2.284
	lognormal	19.959	5.050	250.20	7.750	7.773
	Laplace	0.463	0.308	2.792	1.732	2.463
Critical values (df = 60)	0.10	0.114	0.070	0.790	0.836	1.229
	0.05	0.141	0.083	0.974	0.910	1.311
	0.01	0.208	0.112	1.408	1.062	1.482

The reflectance for each Munsell value is known so the raw data can be translated into the corresponding physical property of the surface. F test done on reflectance values showed that the difference between the targets was significant ($F(1, 120)=85.14$; $p<0.001$). These reflectance values are usually translated into log scale for the purpose of data analysis. In the case of simultaneous contrast the difference is still significant ($F(1, 120)=85.07$; $p<0.001$).

Judging on F ratios alone, it does not appear to be any difference among the conclusions drawn from reflectance vs. log reflectance. However, F ratio calculation is valid only when the distribution is normal otherwise it might lead either to type 1 or type 2 error. Therefore we performed the distribution testing. It was done for the target on the white background only. As can be seen from Figure 3, the results for that target are equal to veridical values, i.e. the white frame ensures most favourable conditions for lightness constancy.

Three distributions were tested: normal, lognormal and Laplace. None of the one-sample tests gave evidence in favour of any of the three hypothesized distributions (table 2). All test-statistics are greater than critical values, for a given degrees of freedom.

At this point it was possible to conclude that there might be some other distribution more applicable for the lightness data. However, we wanted to investigate these particular three distributions due to their theoretical significance and universal usage. Therefore, we hypothesized that a larger sample might fit one of the three distributions. In order to check this more conservative conclusion, log-likelihood ratios between competing distributional hypotheses were calculated. The analysis showed that there is more evidence in favour of normal vs. Laplace ($L = 5.1164$), and normal vs. lognormal ($L = 332.7950$) distribution. Similarly, Laplace vs. lognormal ($L = 327.6847$), favoured Laplace.

Finally, the bootstrapping of the Kolmogorov-Smirnov one-sample test showed that both normal ($p = 0.213$) and Laplace ($p = 0.158$) could be adequate theoretical distributions of the data, while lognormal is proved to be unacceptable ($p = 0.000$). These results are in agreement with the Bayesian inference: of the three theoretical distributions, best model for the participants' judgements is normal.

Based on the analysis we can conclude that the underlying theoretical distribution for reflectance values is most probably normal. Also the double exponential (Laplace) distribution is an adequate fit. However, when the data are transformed into log reflectance values, the normality can no longer be assumed. Consequently the standard repertoire of parametric statistical techniques, and additional statistics suitable for Laplace distribution, should be applied to the reflectance values only.

This work was supported by the Ministry of Science and Environmental Protection in Republic of Serbia, grant number D-149039.

REFERENCES

- Abadie, A. (2002). Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models. *Journal of the American Statistical Association*, 97, 284-292.
- Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (pp 339-351). Cambridge, MA: MIT Press.
- Agostini, T. & Bruno, N. (1996). Lightness contrast in CRT and paper-and-illuminant displays. *Perception & Psychophysics*, 58, 250-258
- Arend, L.E., & Goldstein, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America*, 4, 2281-2285.
- Bressan, P. (2003). A fair test of the effect of a shadow-incompatible luminance gradient on the simultaneous lightness contrast. *Comment. Perception*, 32, 721-723.
- Bressan, P., & Actis-Grosso, R. (2001). Simultaneous lightness contrast with double increments. *Perception*, 30, 889-897.
- Bressan, P., & Actis-Grosso, R. (2006). Simultaneous lightness contrast on plain and articulated backgrounds. *Perception*, 35(4), 445-897.
- Bruno, N., Bernardis, P., & Schirillo, J. (1997). Lightness, equivalent backgrounds, and anchoring. *Perception & Psychophysics*, 59, 643-654.
- Cannon, M. W., & Fullenkamp, S. C. (1991). Spatial interactions in apparent contrast: Inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations. *Vision Research*, 31, 1985-1998.
- Diamond, A. & Jasjeet, S. S. (2005). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. Working Paper. URL: <http://sekhon.polisci.berkeley.edu/papers/GenMatch.pdf>
- Gilchrist, A. (2006). *Seeing Black and White*. Oxford: Oxford University Press.
- Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X. et al. (1999). An anchoring theory of lightness perception. *Psychological Review*, 106, 795-834.
- Gnedenko, B. V. & Gnedenko, D. B. (1982). Laplace Distributions and the Logistic Distribution as Limit Distributions in Probability Theory. *Serdica*, 2, 229-234.
- Katz, D. (1911). Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die individuelle Erfahrung. *Zeitschrift für Psychologie*, 7, 1- 425.
- MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge.
- Puig, P. & Stephens, M. A. (2000). Tests of Fit for the Laplace Distribution, with Applications. *Technometrics*, 42, 417-424.
- Robilotto, R. & Zaidi, Q. (2006). Lightness identification of patterned three-dimensional, real objects. *Journal of Vision*, 6, 18-36
- Salmela V.R. & Laurinen P.I. (2005). Spatial frequency tuning of brightness polarity identification. *Journal of the Optical Society of America A*, 22, 2239-2245.
- Schirillo, J.A., & Shevell, S.K. (1996). Brightness contrast from inhomogeneous surrounds. *Vision Research*, 36, 1783-1796.
- Sivia, D. S. (1996). *Data Analysis: A Bayesian Tutorial*. Oxford University Press, Oxford.
- Winkler, R. L. (2003). *An Introduction to Bayesian Inference and Decision*. Probabilistic Publishing, Sugar Land, TX.
- Zdravkovic, S., Vuong, Q. C., Thornton, I.M. (2005) SLC with static and dynamic backgrounds *Perception (supplement) Vol. 34*