LEUKOS
The Journal of the Illuminating Engineering Society

ISSN: 1550-2724 (Print) 1550-2716 (Online) Journal homepage: https://www.tandfonline.com/loi/ulks20

Derivation of Brightness Scales Using Partition Scaling

Laurens Van de Perre, Wouter R. Ryckaert, Marc Dujardin, Peter Hanselaer & Kevin A.G. Smet

To cite this article: Laurens Van de Perre, Wouter R. Ryckaert, Marc Dujardin, Peter Hanselaer & Kevin A.G. Smet (2019): Derivation of Brightness Scales Using Partition Scaling, LEUKOS, DOI: 10.1080/15502724.2019.1635890

To link to this article: https://doi.org/10.1080/15502724.2019.1635890

Published online: 03 Oct 2019.

LEUKOS
https://doi.org/10.1080/15502724.2019.1635890

Derivation of Brightness Scales Using Partition Scaling

Laurens Van de Perre [a], Wouter R. Ryckaert[a], Marc Dujardin[b], Peter Hanselaer [a], and Kevin A.G. Smet [a]

[a]ESAT/Light&Lighting Laboratory, KU Leuven, Ghent, Belgium; [b]Faculty of Architecture, KU Leuven, Ghent, Belgium

ABSTRACT

To capture the relationship between the magnitude of an optical stimulus (e.g. luminance) and its perceived brightness, a number of psychophysical procedures can be used. One of the oldest procedures capable of capturing such a relationship is partition scaling (PS). A PS procedure constructs an interval scale of a psychological attribute (e.g. brightness) directly from the judgements of observers. It is rarely used in visual research, despite being one of the oldest psychophysical procedures. PS is investigated on its robustness, susceptibility to possible biases and its speed. Partition scaling experiments with simple achromatic discs were set up to obtain brightness scales as a function of luminance. In total four luminance ranges were investigated: three subranges (low-, mid- and high-range) equally divided from 5 to 175 cd/m^2 and one overlapping full-range. Results show that observers had difficulties in accurately estimating the mid- and high-ranges, because the brightness of the two anchor discs was too similar. However, perceptual brightness scales could be obtained for the full- and low-range. To reduce cumulative errors and increase observer accuracy, an improved PS method was implemented. Results show that PS is a valid, rapid psychophysical procedure capable of capturing a brightness scale without the need of prior knowledge of the psychophysical brightness function.

ARTICLE HISTORY
Received 15 January 2019
Revised 18 June 2019
Accepted 21 June 2019

KEYWORDS
Partition scaling; psychophysical methods; brightness perception; visual perception

1. Introduction

Establishing the relationship between the luminance of a stimulus and its perceived brightness is crucial in several scientific domains, such as colour appearance modelling, lighting design, evaluating visual comfort criteria and lighting research in general. Several psychophysical procedures exist to capture the relationship between the magnitude of a (psycho)physical stimulus and the magnitude of its corresponding percept (henceforth referred to as "perceived magnitude").

Kingdom and Prins (2016) presented a classification scheme including an overview of a variety of these psychophysical procedures. In this classification scheme two main types of psychophysical procedures exist, namely performance-based and appearance-based procedures. Performance-based procedures measure the performance of an observer; i.e. how good an observer is at a particular visual task. Appearance-based procedures are generally used to acquire the perceived magnitude of a physical stimulus and consists of two major categories, i.e. the matching and scaling procedures. Matching procedures aim to measure the point of subjective equality between stimuli whereas scaling procedures aim to find the relationship between the physical magnitudes and perceived magnitudes. Partition scaling is an appearance based scaling procedure.

1.1. Partition scaling

Partition scaling (PS) construct an interval scale of a psychological attribute (e.g. brightness) directly from the judgments of an observer without any translation from perceptual to numerical representation. It is rarely used in visual research, despite being one of the oldest psychophysical procedures. Most PS methods are based on a bisection task, wherein an observer is shown two discs of different brightness and is asked to adjust the brightness of a third disc such that two equal-appearing perceptual intervals are produced. These two equal-perceptual intervals obtained using the bisection method can be further progressively bisected in smaller intervals. Care should be taken to minimize cumulative errors, as each subsequent interval is dependent on the previously set interval. The

CONTACT Laurens Van de Perre laurens.vandeperre@kuleuven.be ESAT/Light&Lighting Laboratory, KU Leuven, Ghent 9000, Belgium.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ULKS.

output of this progressive solution method is an equisection scale. According to Stevens (1975), bisection is the earliest version of equisection scaling and was invented by Plateau in the 1850's. Plateau asked eight painters to mix a grey mid-way between a given white and black and reported that the eight greys turned out to be nearly all the same (Laming and Donald 1997). The term equisection is generally used when the observer sections multiple equal intervals whereas in bisection only two equal perceptual intervals are created (Gescheider 1997).

Instead of progressively building the equisection scale, as used in the progressive solution method, the scale can also be constructed simultaneously. For example, an observer is shown two anchor stimuli (e.g. a dim and a bright disc) and is required to adjust the perceived brightness of $n - 1$ stimuli, resulting in n equidistant sections. This latter method has been termed simultaneous solution by Gescheider (1997) and multipartition scaling by Kingdom and Prins (2016). A disadvantage of this technique is that adjusting a single disc could require the adjustment of previously adjusted discs, resulting in a number of iterations before achieving a satisfactory result. A well-known experiment that used this technique was conducted by Whittle (1992). Observers were presented with an arrangement of several discs in the form of a spiral on a uniform grey background on a display. The first and last discs were anchors set to the lowest and highest luminance values available on the display. The observers' task was to adjust the brightness values of the remaining discs to create an equal-interval scale.

1.2. Biases

Most psychophysical procedures suffer from several biases. For example, centring bias can occur as observers tend to centre their range of response on the range of presented stimuli (Poulton 1979). Range effects occur when the outcome of the perceptual sensation is influenced by the selected stimulus range (Teghtsoonian and Teghtsoonian 1978).

Bisection, as well as equisection methods has been shown to be affected by order effects or sequential and hysteresis effects. An order or sequential effect is an effect wherein the current bisected value is dependent on the order of presentation. More specifically, when this effect occurs from sequentially presenting the end points of the current interval in ascending or descending order of intensity, it is commonly referred as "hysteresis" effects. This effect has been found in psychophysical experiments on brightness, loudness and heaviness (Stevens 1957). For example, in a sound intensity experiment, the observers' bisection estimate was found to be different for ascending or descending orders in tones (Garner 1954). A similar effect was observed in a brightness experiment (Stevens 1961; Stevens and Stevens 1960). The "hysteresis" effect can be partially counterbalanced by presenting the end points simultaneously, however sometimes this is not feasible, such as in a sound intensity experiment (Stevens 1975).

In this paper, PS is investigated on its robustness, susceptibility to possible biases and its speed using a progressive solution technique for brightness scaling. An improved PS method is presented reducing cumulative errors and increasing observer accuracy.

2. Methodology

2.1. Experiment setup

Neutral discs varying only in luminance were presented on a calibrated colour monitor (ColorEdge CG246) with a refresh rate of 60 Hz, a resolution of 1,920 × 1,200 pixels and a colour depth of 10 bits per channel. The background on the monitor surrounding the discs was black (< 0.5 cd/m^2) and the room was completely darkened during the experiments (unrelated stimuli), apart from the observer monitor and experiment supervisor monitor (fully dimmed). Figure 1 shows a picture of the experimental room on the left and a schematic layout of the room on the right. The monitor was calibrated up to approximately 180 cd/m^2 (CIE 1931 2° observer) using a colorimetric imaging camera (TechnoTeam LMK5-5 Color) to ensure accurate presentation of the luminance of the discs, that could be displayed on the left, right and middle of the monitor. The monitor was calibrated at the start of each experiment day by measuring all possible (2^{10}) grey values for each disc position. For each location of the discs, a gain-gamma-offset (GGO) model was fitted (Katoh et al. 2001) to the calibration measurements. Luminance accuracy was checked for each disc location and was within 1% of the requested values. Observers were

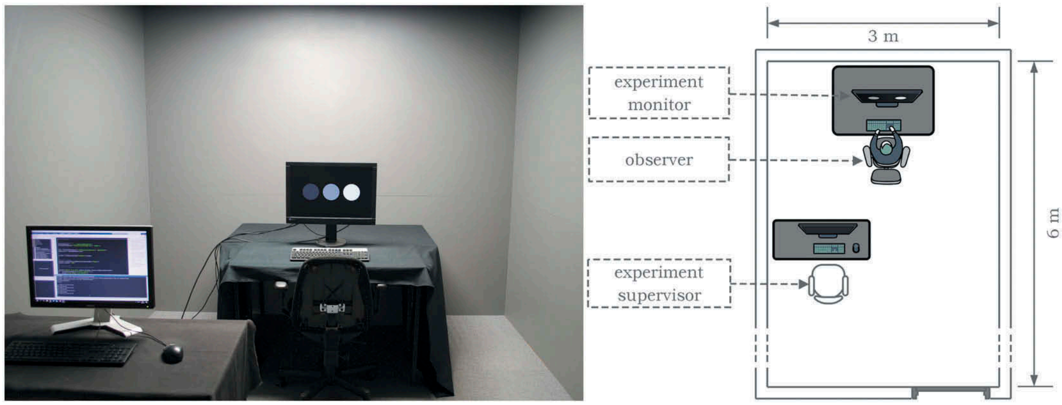

Fig. 1. Left: Image of experiment room. Note that during the experiment the stimuli and the supervisor monitor were the only light source in the room. In the image, the ambient lighting in the room was on for visual clarity. Right: A schematic layout of the experiment room.

seated at a distance of roughly 60 cm from the monitor, resulting in a field of view for the monitor of approximately 40° and of about 10° for the central disc. The experiment was programmed in MATLAB R2018a and stimuli were generated using the Psychophysics toolbox (version 3.0.14) (Brainard 1997). Observer gave their response using a keyboard.

2.2. Experiments

Three partition scaling experiments were conducted to determine brightness scales. These PS experiments were conducted as part of a series of experiments investigating several psychophysical procedures to produce brightness scales. In this paper only the PS experiments and their results are presented. A follow-up paper will compare the PS results with those obtained using the commonly adopted magnitude estimation method.

The PS experiments were divided over three sessions. The first session consisted of three experiments each testing a different psychophysical method with a random order for each observer. One of the three experiments was a "standard" PS experiment where the lowest luminance anchor point was always presented on the left on the monitor. A standard PS experiment with the low luminance anchor on the right was performed separately in session two. An improved PS experiment was conducted in session three.

Fifteen observers – six females and nine males – with ages ranging from 24 to 30 years (average 26.3 years) participated in the first two sessions. Thirteen observers – five female and eight male – between 24 and 41 years old (average 28.5 years) participated in session three. Observers participating in session one also completed session two. Ten observers participated in all three sessions. There were on average 11 and 26 days between session one and two, and session two and three, respectively. Prior to the experiment, participants gave a written informed consent. The study was conducted in agreement with the social and societal ethics committee (SMEC) of KU Leuven. External participants were compensated at a rate of 10 €/hour.

Four luminance ranges were investigated. A *full-range*, from 5 cd/m^2 to 175 cd/m^2 and three subranges: a *low-range*, from 5 cd/m^2 to 82.3 cd/m^2; a *mid-range*, from 51.4 cd/m^2 to 128.6 cd/m^2 and a *high-range* from 97.7 cd/m^2 to 175 cd/m^2. Each subrange had a 40% luminance overlap with its neighbouring range(s) and spanned 77.3 cd/m^2. Although using a logarithmic or exponential function to subdivide the full range, would result in somewhat more equal perceptual stimulus ranges, a linear subdivision was chosen to avoid any prior knowledge of the brightness scale that could potentially influence the results.

2.2.1. Standard PS method

In each PS experiment, an observer is shown two discs on the left and right of the display, each with a specific "anchor" luminance, and a third central disc. A PS task consists of bisecting the brightness difference between the two anchor discs into two perceptually equal parts by adjusting the luminance of the central disc.

Two new sets of anchor points (luminance values) are thereby obtained, i.e. left-centre and centre-right, which can each be further bisected. The process can then be repeated by creating progressively smaller luminance intervals between anchor points as the level of subdivision increases, as shown in Fig. 2. At subdivision level 0, the luminance values of the anchor points (relative brightness step 1 and 9, denoted as "F") are fixed to that of the endpoints of a luminance range. At subdivision level 1, two luminance intervals are created (sets of anchor points, i.e. relative brightness steps 1 & 5 and 5 & 9). In subdivision level 2 and 3, respectively four and eight luminance intervals are created. A completed PS with three subdivision levels, results in eight equal perceived brightness intervals (partition scale) and nine luminance values (two for the two fixed endpoints and seven for the observer adjusted discs).

There are four ways of running a PS. They depend on the position (left or right) of the minimum and maximum luminance of the anchor points and on starting luminance (high or low) of the central disc. The central disc starting luminance was either dark/low (< 0.5 cd/m^2) or bright/high (~180 cd/m^2). The four possible ways (left/right, bright/dark) of running a PS were split over the two sessions, each consisting of all four luminance ranges. In session one, the minimum anchor point luminance was always to the left. In session two it was always to the right of the display. For both sessions, the starting luminance of the central disc could either be bright or dark and was fully counterbalanced. The two possible starting conditions (dark/bright) of the central disc and the four luminance ranges were pooled per subdivision level and run in a random order, i.e. all possible bisections for a subdivision level were completed in random order before moving to the next level. The PS was run up to and including subdivision level 3.

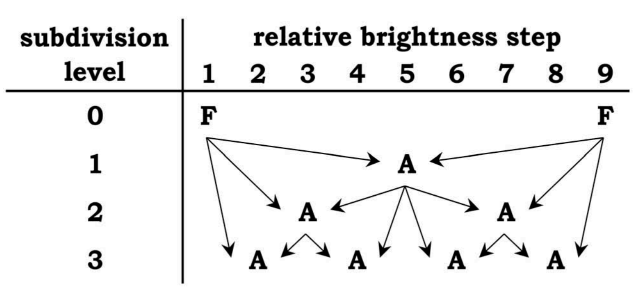

Fig. 2. Schematic overview of a partition scaling experiment with eight equal perceptual brightness intervals and three sublevels. The starting anchor discs (fixed anchor disc) are denoted as "F". An adjustable disc is denoted with "A" and the corresponding anchor discs for that bisection task are denoted with two arrows going from the previous sublevel(s) to the current task.

Before the start of each standard PS experiment (session 1 and 2), the observer was given brief verbal instructions (see appendix) followed by a short PS training with two subdivision levels for the full luminance range. In total each observer completed 112 bisection tasks for the two sessions (7 adjustable discs per complete PS times 4 luminance ranges times 4 possible ways running the PS). This resulted in a total of 16 brightness interval scales (4 luminance ranges times 4 possible ways of running the PS). A dark screen was shown for one second between bisection tasks and observers were not imposed with a time limit to complete a bisection task.

2.2.2. *Improved PS method*

In the improved PS experiment, the four possible ways (left/right, bright/dark) of running a PS experiment were combined per subdivision level and were run in a random order, i.e. all possible bisections for a subdivision level were completed before moving to the next subdivision level. After a subdivision level is completed, each anchor point for the next subdivision level was determined as the arithmetic mean luminance of all four (left/right, bright/dark) observer adjusted luminance values. This improvement aims to minimize the possible accumulation of errors as the bisection progresses. The improved PS was run up to and including subdivision level 4 – resulting in 15 adjusted discs –, and only for the full luminance range. A schematic overview of the improved PS is shown in Fig. 3.

The improved PS experiment was conducted in session 3. Before the start of the main experiment, the observer was given verbal instructions (see Appendix) and a short PS training. The training consisted of completing two subdivision levels for the full luminance range. Four brightness scales were obtained for the full range, one for each way of running the PS. In total, each observer made 60

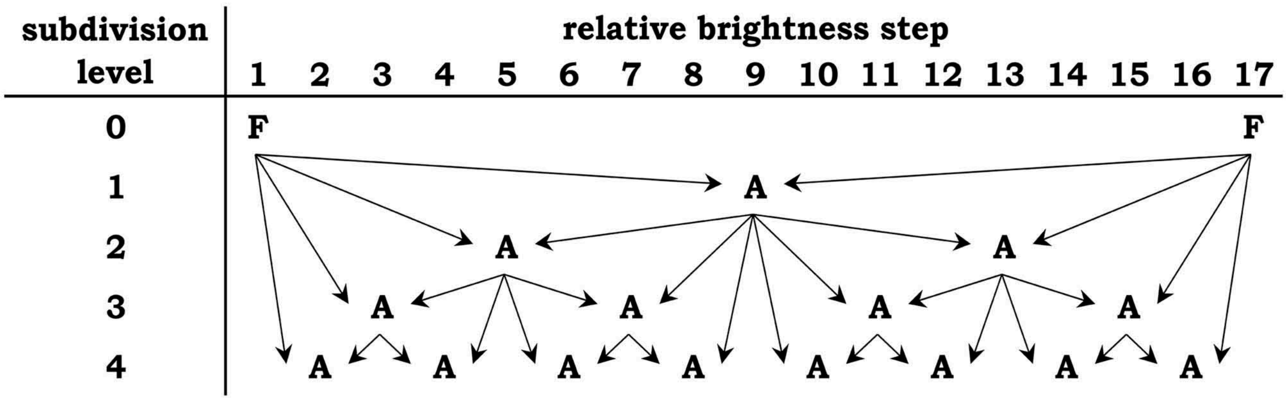

Fig. 3. Schematic overview of a partition scaling experiment with 16 equal perceptual brightness intervals and 4 sublevels. The starting anchor discs (fixed anchor discs) are denoted as "F" and adjustable discs are denoted with "A".

bisections (15 discs adjustments per PS times 4 ways to run a PS).

2.3. Brightness scale models

Several psychophysical models and formulas exist that define a relationship between the magnitude of a physical stimulus and its perceived magnitude. One of the most frequently used is Steven's power law (Stevens 1975) and can be used to model the luminance-brightness relationship. It states that the brightness (Q) grows as a function of luminance (L) as follows:

$$Q = aL^b \tag{1}$$

Where a is a scaling factor and b the exponent of the power function. Commonly reported values for b in literature are between 1/3 and ½ (Stevens 1957, 1975; Withouck et al. 2015).

The output of a PS experiment is an interval scale, which can be rescaled using a simple linear function. The brightness power function (1) can therefore be transformed to a brightness function for PS:

$$Q = aL^b + c \tag{2}$$

Where a is a scaling factor, b is the exponent of the power function and c is a constant offset. The brightness step Q corresponds to the adjusted luminance of a disc.

However, instead of only considering the brightness step, a more general approach would be to consider the luminance of the anchor points and the adjusted disc from the PS task. To briefly summarize the observers task; the observer is presented with two discs on the left and right, each with a specific luminance, L_{left} and L_{right}, respectively. The observer adjusts the luminance of the middle disc (L_{mid}) such that two perceptually equal brightness intervals are created between the left & middle and the middle & right discs. The brightness perception of the left (Q_{left}), middle (Q_{mid}) and right (Q_{right}) discs can thus be written as follows:

$$\begin{aligned} Q_{mid} - Q_{left} &= Q_{right} - Q_{mid} \\ Q_{mid} &= \left(Q_{left} + Q_{right}\right)/2 \end{aligned} \tag{3}$$

The latter can be rewritten in terms of luminance values using the power law (1):

$$L_{mid} = \sqrt[b]{\frac{{L_{left}}^b + {L_{right}}^b}{2}} \tag{4}$$

The brightness function (4) estimates the luminance of the middle disc using the luminance of both anchor points and a brightness exponent b. This function is later on used in the results and is referred to as the general model.

3. Results & Analysis

The results of the standard and improved PS method are discussed first followed by a comparison between both methods.

3.1. Standard PS method

First, the observer accuracy is discussed, followed by the results of the separate and pooled luminance ranges for the average observer. Afterwards the general model approach is fitted to the individual observer data. Finally, a repeated measure ANOVA is conducted, investigating several (possible) biases.

3.1.1. Observer accuracy

In a brightness bisection task the observer is asked to adjust the brightness of the middle disc such that two equal perceptual brightness intervals are created. Brightness for simple neutral stimuli is expected to follow a monotonic increasing function as a function of stimulus luminance (power law). The luminance of the adjustable disc (L_{mid}) should therefore always have a value between the luminance of both anchor discs (L_{left} and L_{right}). However, when the luminance difference between the two anchor discs is too small, of the order of a just-noticeable-difference (JND), an observer may have difficulties setting a luminance in-between the two anchors. An observer bisection error is defined as the percentage of bisection tasks where an observer adjusts the luminance level (L_{mid}) outside the luminance anchor discs interval: $\left]L_{left}; L_{right}\right[$. Table 1 shows the average observer bisection error (in percentage) for each luminance range per individual subdivision level (1, 2 and 3) and all subdivision levels combined (all), per anchor order and start luminance of the adjustable disc. The anchor order L < R means $L_{left} < L_{right}$ and vice versa.

Results show that each observer could perform the first bisection task (subdivision level 1) of every range without any bisection error. The average observer bisection error increased with subdivision level and with luminance range from full- to low- to high-range, indicating that observers had increasing difficulty executing bisection tasks for smaller anchor luminance intervals and/or higher luminance values. In the full- & low-range only 2% and 4% bisection errors occurred. For the mid- & high-range the error increased to respectively 19% and 31%, indicating these two ranges were possibly too difficult to bisect at higher subdivision levels, which is consistent with observer feedback.

The *R* language (R Core Team 2019) and package *lme4* (Bates et al. 2015) was used to perform a generalized linear mixed-effects model (GLMM) analysis on the observer bisection errors. An observer either made a mistake (1) or not (0) per bisection task. A binomial distribution with a logit link function was therefore used for the response variable. Maximum likelihood was used as estimation method for the model. The fixed effects in the model included the range and subdivision level (without interaction term) along with anchor order and start value (with interaction term). The only random effects included were the observer intercepts. No presence of overdispersion was found ($\chi^2(711) = 318.634, p = 1$). Visual inspection of residual plots did not reveal any sign of heteroscedasticity (i.e. non-homogeneity of the variance) among the residuals. The random intercepts were tested on normality using the Shapiro-Wilk's method, normality is assumed ($p = 0.432$). Significance of fixed effects were tested using the Wald Chisquare test (Table 2). The results show that only the luminance range and subdivision level is significant, as expected. Increasing the luminance range results in higher luminance values for the anchors in a bisection task, whereas increasing the subdivision level results in smaller luminance intervals between the anchors. Both effects result in an increase of bisection errors. There was no significant effect of anchor order position and starting luminance value (including interaction term) on the bisection errors.

The duration for each bisection task per observer was also recorded during the experiments. On average, observers completed a bisection task in 26.12 seconds.

Table 1. Average observer bisection error [%] structured by ranges and subdivision level per anchor order position and starting value of the adjustable disc. The row below the name of the luminance ranges denotes the subdivision level, "all" represents the combination of all subdivision levels bisection errors and 1 to 3 corresponding to the individual subdivision level bisection error. The last row shows the mean bisection errors per range and subdivision level.

		Full-range				Low-range				Mid-range				High-range			
Anchor order	Start value	All	1	2	3	All	1	2	3	All	1	2	3	All	1	2	3
L < R	Dark	1.0	0	0	1.7	3.8	0	0	6.7	15.2	0	10.0	21.7	30.5	0	23.3	41.7
L < R	Bright	2.9	0	0	5.0	1.9	0	0	3.3	20.0	0	6.7	31.7	31.4	0	23.3	43.3
L > R	Dark	1.9	0	0	3.3	4.8	0	0	8.3	21.9	0	3.3	36.7	31.4	0	13.3	48.3
L > R	Bright	2.9	0	0	5.0	5.7	0	0	10.0	18.1	0	3.3	30.0	32.4	0	26.7	43.3
	Mean	2.1	0	0	3.8	4.0	0	0	7.1	18.8	0	5.8	30.0	31.4	0	21.7	44.2

Table 2. Results of Wald Chisquare test for fixed effects. The columns show from left to right; the fixed effect, Chisquare value, degrees of freedom (DoF) and the p-value.

	χ^2	DoF	p
Range	151.183	3	<0.001
Subdivision level	86.803	1	<0.001
Anchor order	1.319	1	0.251
Start value	0.486	1	0.486
Anchor order : Start value	0.336	1	0.562

Results show a slight increase in duration from full- to low- to high-range with 24.9, 25.7, 26.1 and 27.7 seconds, respectively; with longer duration possibly indicating the increased difficulty in bisecting the higher luminance ranges.

3.1.2. Brightness perception: individual luminance ranges

Each observer completed each luminance (sub)range in all four possible ways (anchor order: left/right and starting luminance: bright/dark). No major differences of bisection errors between anchor order and starting luminance of the adjustable disc were found. Therefore, per luminance range and for each observer, the adjusted luminance values for both anchor orders and starting luminance values are combined using the arithmetic mean. These individual observer averaged adjusted luminance values are then combined using the arithmetic mean over all observers, resulting in average observer adjusted luminance values per luminance range (shown in Fig. 4). A relative brightness step can be seen as a brightness value on an equidistant relative brightness scale, since the outcome of a PS experiment is an equal perceptual interval scale. The error bars represent standard errors on the arithmetic mean over all observers. The fixed anchor points (relative brightness step 1 and 9) have no standard errors since they were fixed values depending on the luminance range. The results indicate that as the luminance increases the standard error also increases.

In Fig. 4, the brightness models that were fitted for each luminance range using (2), are shown as coloured dotted lines. The parameters a, b and c, their 95% confidence intervals (CI) and the coefficients of determination (R^2) are shown in Table 3. All fits show very high R^2 values, even for the mid-

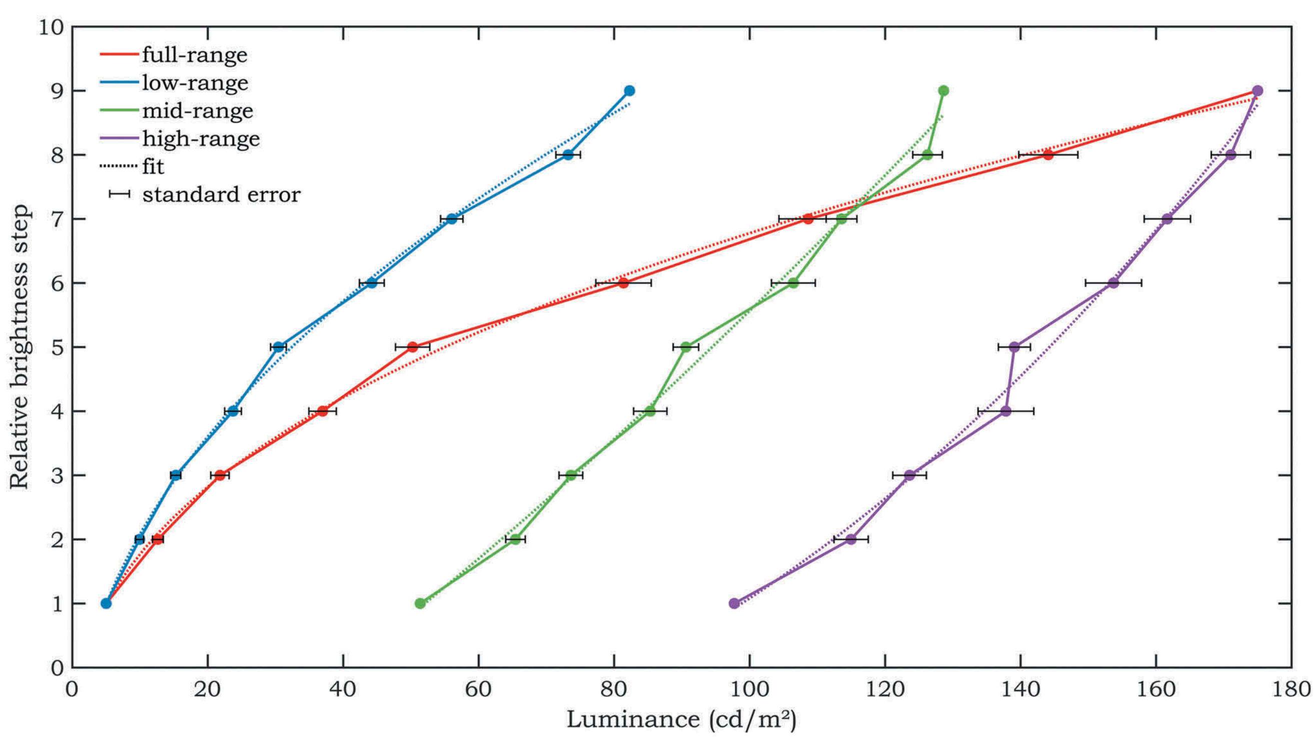

Fig. 4. Average observer adjusted luminance values per relative brightness step are shown for each luminance (sub)range. The full-, low-, mid- and high-luminance ranges are highlighted in red, blue, green and purple, respectively. The error bars are standard errors on the arithmetic mean per observer. Each luminance range is fitted with a power law equation denoted by a coloured dotted line.

Table 3. Fitting results per range with average observer adjusted luminance value as input and relative brightness step as output. The 95% confidence intervals (*CI*) and parameters *a, b* and *c* are shown. The coefficient of determination (R^2) of the model is shown in the last column.

	Parameter *a*		Parameter *b*		Parameter *c*		
	Value	*CI*	*Value*	*CI*	*Value*	*CI*	R^2
Full-range	1.268	[0.539, 1.997]	0.406	[0.316, 0.497]	−1.463	[−2.739, −0.187]	0.998
Low-range	1.306	[0.330, 2.281]	0.475	[0.339, 0.611]	−1.816	[−3.566, −0.066]	0.997
Mid-range	0.023	[−0.070, 0.115]	1.275	[0.507, 2.044]	−2.511	[−6.650, 1.628]	0.991
High-range	0.000	[−0.001, 0.001]	2.237	[0.983, 3.490]	−2.002	[−5.552, 1.548]	0.990

& high-ranges, despite their higher bisection errors for the average observer. This illustrates that high R^2 for average observer data is not always a good indicator. If there is no range effect present, all *b* values should be similar. The full- & low-range show a similar *b* value and the confidence intervals overlap. However, both mid- & high-range *b* values are very high and differ substantially from the full- & low-range, as indicated by their (partially) non-overlapping confidence intervals. This is in agreement with the high average observer bisection errors for these ranges and reinforces the statement that the mid- & high-range may not be reliable. This was also confirmed by observers stating they had difficulty bisecting these ranges. Again, high R^2 values of fits on average observer data are illustrated to not be an adequate indicator for successful task completion.

3.1.3. Brightness perception: pooled luminance ranges

To pool the PS data of all ranges, they were transformed to a common scale using a linear transformation to ensure the interval properties of a PS scale are kept intact. A similar method was used by (Gescheider 1997; Stevens and Volkmann 1940) to combine three overlapping equisection frequency scales on auditory pitch. First, for each luminance range a brightness model is fitted with (2) (fitting results shown in Table 3), using the average observer adjusted luminance values as input and the relative brightness steps as output (shown in Fig. 4). Second,

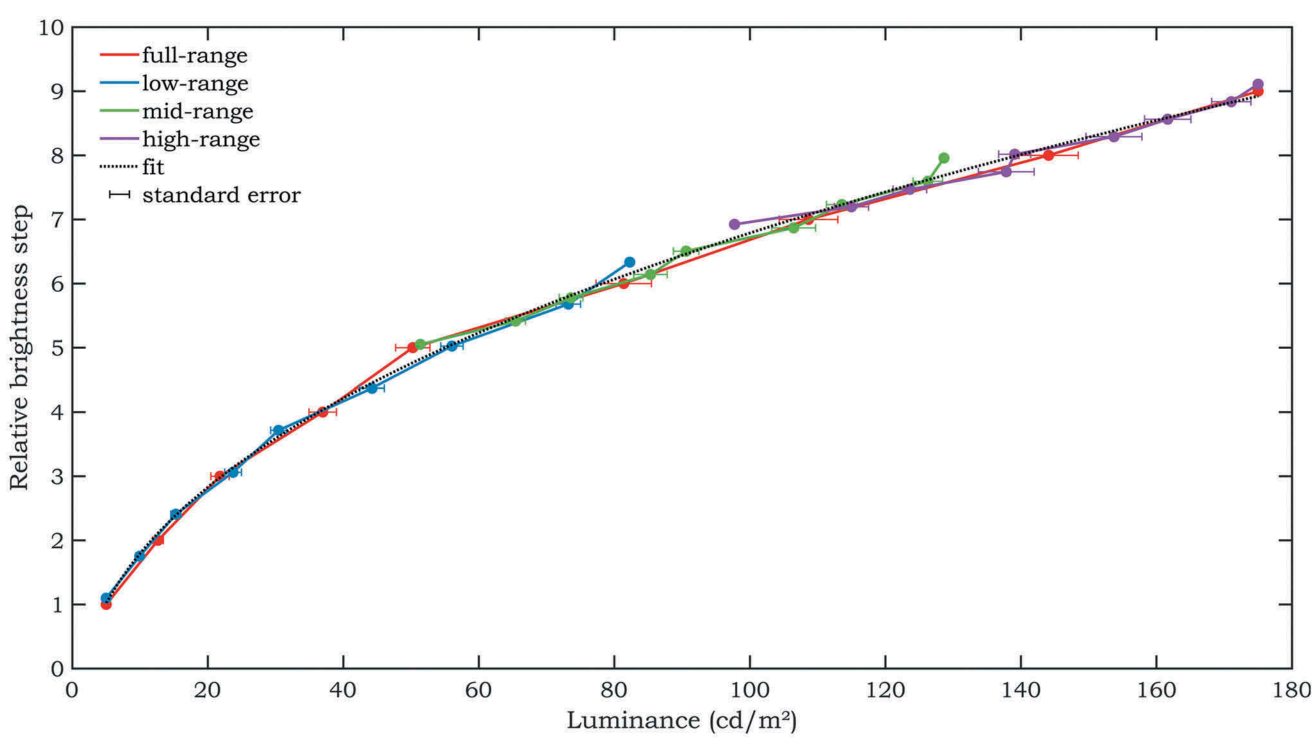

Fig. 5. Adjusted luminance values of the average observer for each (sub)range rescaled to the full-range brightness scale. The luminance ranges: full-, low-, mid- and high-range are highlighted in red, blue, green and purple, respectively. The error bars are standard errors on the arithmetic mean over all observers. The brightness model fitted to the pooled data is plotted as a black dotted line.

a linear scaling function is fitted to the predicted relative brightness step – obtained using the brightness models – for the shared luminance interval for each pair of overlapping luminance ranges. These linear functions (scaling factors with offsets) are then used to transform the brightness data of the luminance subranges to that of the full-range. The rescaled data is shown in Fig. 5 using the full-range as common scale. The error bars represent standard errors on the arithmetic mean of the adjusted luminance values over all observers.

Finally, a brightness model is fitted to the pooled rescaled data, shown as a black dotted line, with R^2 of 0.997 and parameters $a = 1.162$, $b = 0.420$ and $c = -1.263$. The exponent b (CI: [0.380, 0.461]) is close to that of the separate fits of the full- & low-range (0.406 and 0.475), but as expected, is substantially different from the b values of the mid- & high-range (1.275 and 2.237). Based on Fig. 5, there is no clear indication for a range effect as each luminance range overlaps quite well. Results show that as the luminance increases the standard error increases.

Table 4. Exponent b, 95% CI and R^2 for fits to the individual observer data using the general brightness model (4) for each of the individual luminance ranges and several pooled ranges. The first 4 rows show fits for separate ranges and the last three rows show fits for pooled ranges.

	Parameter b		
(pooled) ranges	*Value*	*CI*	R^2
Full-range	0.403	[0.363, 0.444]	0.942
Low-range	0.490	[0.436, 0.543]	0.936
Mid-range	1.311	[1.060, 1.562]	0.851
High-range	1.825	[1.306, 2.345]	0.758
Full- & low-range	0.414	[0.384, 0.444]	0.948
All ranges	0.437	[0.403, 0.470]	0.958
Low- to high-range	0.644	[0.558, 0.730]	0.961

3.1.4. Brightness perception: general model

The previous rescaling and pooling method is only applicable to PS data and is dependent on the presence of overlapping luminance ranges. Using (4) does not require several overlapping luminance ranges and avoids the need for a prior rescaling of the individual ranges. It predicts the adjusted luminance (L_{mid}) using the luminance of both anchor points and a brightness exponent b. Figure 6 shows the unique adjusted

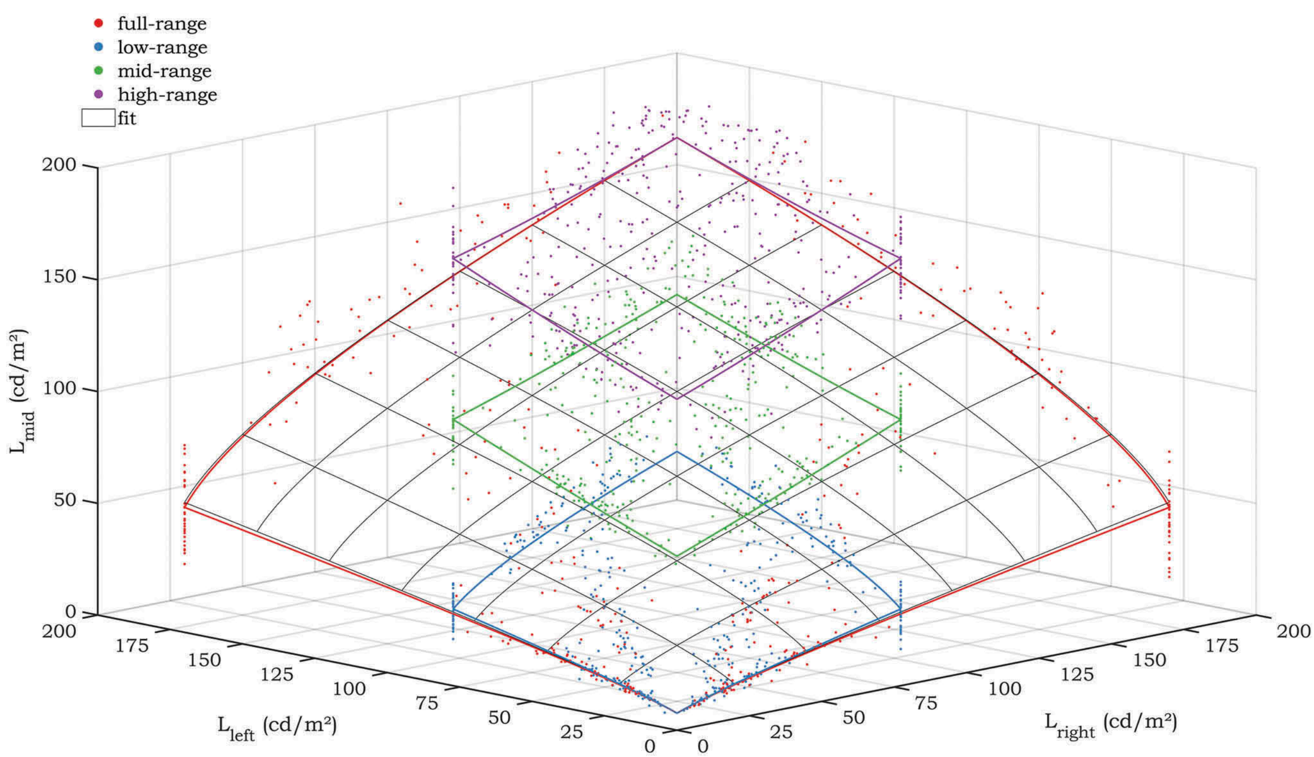

Fig. 6. Adjusted luminance values of the individual observers (dots) and separate fitted general models (surface fits shown as coloured borders) for the four luminance ranges. Data for the full-, low-, mid- and high range are plotted in red, blue, green and purple, respectively. The black grid represents the fitted general model of all ranges pooled.

luminance values per bisection for each observer as coloured dots. The general model fitted to each of the luminance ranges and to the pooled ranges are shown as coloured borders and a black grid, respectively. A large spread is visible for the initial adjusted luminance values for each range. These initial adjusted luminance values could predominantly influence the fitting results. The results also show more spread as the luminance of the anchor points increases, this is also consistent with previous results.

The exponent b, with 95% confidence intervals (CI), and the coefficients of determination (R^2) for model fits to the data of the individual ranges and several pooled ranges are shown in Table 4.

Comparing b values for the individual luminance ranges obtained by fitting the average observer data using (2) with those obtained by fitting the individual observer data using the general model (4) shows that the mid- & high-range have slightly different b values. However, overall no substantial differences were found, as confirmed by their overlapping confidence intervals. As expected, the models fitted to the individual observer data have substantially lower R^2 values (see Table 4) than those fitted to the averaged observer data (see Table 3). Of the general models fitted to the individual observer data of the separate ranges (Table 4, row 1–4), the R^2 value for the full-range was highest, whereas the R^2 for the subranges show a systematic decrease from low- to high-range. The latter suggests observers become more and more inconsistent with increasing luminance range. This is consistent with observer feedback and average observer bisection errors.

The model for the pooled subranges has a high R^2 (0.961), with a b value of 0.644. It is substantially different from the exponent of the full-range (0.403), suggestive of a range effect (despite no clear effect can be observed from Fig. 5). However, the results for the subranges could have been biased by the difficulty observers experienced when bisecting the mid- & high-ranges. The results show that as the observer inaccuracy increases, such as the case for the mid- & high-range data, the b value increases. The b values of the full-range (0.403) and low-range (0.490) model fits show a moderate difference. The slightly overlapping confidence intervals suggest there is no substantial range bias (as confirmed by a statistical test below). A comparison of the b values between the pooled full- & low-range (b = 0.414) and the individual ranges shows no indication for any range bias, except for the mid- & high-ranges (b = 1.311 and b = 1.825, respectively).

3.1.5. Position bias, luminance starting value and range bias

Determining the presence of a range, position and luminance starting value bias on the luminance values of individual adjusted discs is difficult since each subsequent adjusted value is dependent on the luminance of the anchor points determined in the previous bisection tasks. The average observer bisection errors are a first indicator. When no biases are present, the average errors are similar for both anchor order and starting value of the adjustable disc. However, a range bias could still be present despite similar average observer bisection errors over several luminance ranges. Previous analyses did not indicate the presence of a range bias, except for the mid- & high-ranges. However, the difficulty experienced by observers when bisecting them makes them unreliable. Therefore, to determine the impact of position, luminance starting value and luminance range on the results, a factorial repeated measures ANOVA is conducted on the b values obtained by fitting the general model (4) to the luminance adjustments of each individual observer. In total there were three factors: luminance range, position (anchor order) and start value, with 4, 2 and 2 levels, respectively.

Normality of b values was confirmed ($p > 0.05$) for each unique condition (16 in total) using Shapiro-Wilk's method. Mauchly's test indicated that the assumption of sphericity was violated for the main effect of (luminance) range, $\chi^2(5) = 61.070, p < 0.01$. The Greenhouse-Geisser corrected test is therefore reported ($\varepsilon = 0.393$). Sphericity was not violated for the 2-level main effects of anchor order and starting value. No significant effects were found for anchor order $F(1, 14) = 0.198, p = 0.663$ and starting value $F(1, 14) = 0.768, p = 0.395$, confirming the lack of position and luminance starting value biases on the exponent values b. This is also consistent with results of the GLMM analysis on bisection errors. The main effect of luminance range was found to be significant $F(1.179, 16.500) = 21.477, p < 0.01$.

Contrast tests revealed that each pair of ranges was found to have significantly ($p < 0.05$) different b values. The full- & low-range pair had $F(1, 14) = 5.278, p = 0.038, r = 0.523$ while all other pairs had $p < 0.01$. While contrast tests provide effect size and significance, they do not adjust for multiple comparisons. Therefore, a post hoc pairwise comparisons test was calculated using the Bonferroni adjustment for multiple comparisons. Results show that the b values were significantly different for all pairs, except the full- & low-range ($p = 0.225$). This indicates that there was a range bias between all pairs except the full- & low-range.

3.2. Improved PS method

The improved PS method was conducted in the last (3rd) session for only the full luminance range with four subdivision levels for all four ways to run a PS. The results of the improved method are presented in a similar structure as the standard method.

3.2.1. Observer accuracy

Table 5 shows the average observer bisection error [%] for the full-range in the improved method. Results are subdivided by subdivision levels (i.e. "all" and 1–4), anchor order position (L < R or L > R) and luminance starting value (dark or bright). Results show that observers could bisect up to and including subdivision level 3 (first 7 bisection tasks) without any bisection errors, except for one bisection error by a single observer in the third level. Average observer bisection error for the 3rd subdivision level (over the four possible ways of running the PS) is therefore less than 0.5%, compared to 3.8% for the standard PS method with the full-range. This shows that the improved PS method increases observer accuracy and decreases possible accumulation errors. The bisection errors for the 4th subdivision level in Table 5, show a slight impact of anchor order position, with lower errors for the L < R anchor order.

Table 5. The average observer bisection error [%] of the improved PS for the full-range. Bisection errors are subdivided by subdivision level, anchor order position and luminance starting value. The 2nd header row denotes the subdivision levels (1–4), "all" represents the combination of all four subdivision levels. The last row shows the mean bisection errors per subdivision level.

		Full-range				
Anchor order	Start value	All	1	2	3	4
L < R	Dark	6.7	0	0	0	12.5
L < R	Bright	6.2	0	0	0	11.5
L > R	Dark	8.2	0	0	0	15.4
L > R	Bright	9.7	0	0	1.9	17.3
	Mean	7.7	0	0	0.5	14.2

A GLMM is used to analyse the observer bisection errors, similar to analysis performed for the standard method (see section 3.1.1). The same fixed effects were used as before, except range was not included as for the improved PS method, data was only collected for the full range. No presence of overdispersion was found ($\chi^2(202) = 61.037, p = 1$). Visual inspection of residual plots did not reveal any sign of heteroscedasticity among the residuals. The random intercepts were tested on normality using Shapiro-Wilk's method, normality is assumed ($p = 0.110$). Of all fixed effects (including interaction term) only the subdivision level was found to be significant ($\chi^2(1) = 13.200, p \leq 0.001$).

On average, observers completed a bisection task in 26.10 seconds, which is slightly higher compared to the average duration of the full-range of the standard method.

3.2.2. Brightness perception: Average observer

The adjusted luminance values for both anchor order and starting luminance value are combined using the arithmetic mean, for each observer. These individual observer averaged adjusted luminance values are then combined using the arithmetic mean over all observers, resulting in average observer adjusted luminance values per luminance range (shown in Fig. 7). Results show that as the luminance increases, the standard error also gradually increases. The brightness model fitted using (2) is shown as a black dotted line. The model fit had a very high R^2 of 0.999. The model parameters have the following values: $a = 2.334$, $b = 0.423$ and $c = -3.753$. The confidence interval of parameter b is $[0.381, 0.465]$. The confidence interval and b value is nearly the same as the b value ($b = 0.420$) of the model fitted to the rescaled, pooled average observer data and the b value ($b = 0.406$) of the model fitted to the full-range average observer data, both obtained using the standard method (see section 3.1.3). This indicates that the improved and standard method obtain similar end results.

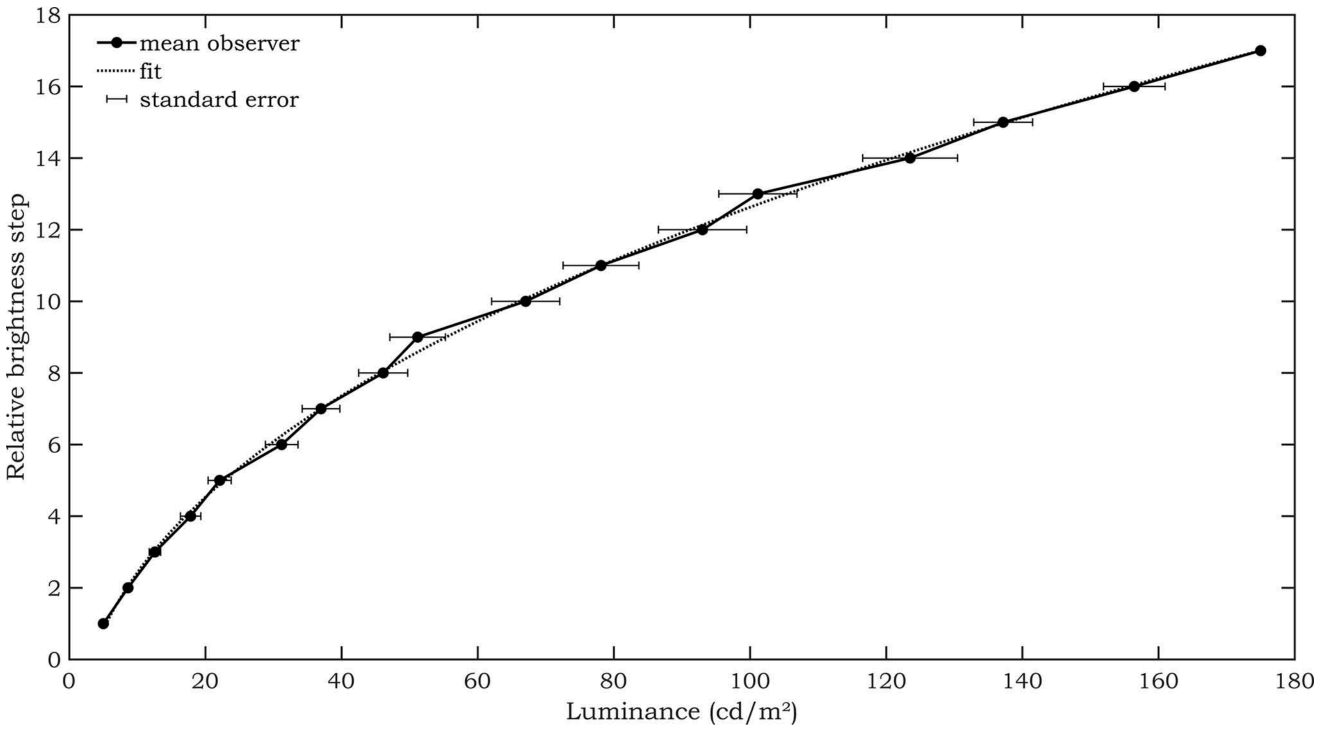

Fig. 7. Adjusted luminance values of the average observer for the improved PS method for the full-range. The error bars are standard errors on the arithmetic mean over all observers. A brightness model is fitted to the data shown as a black dotted line.

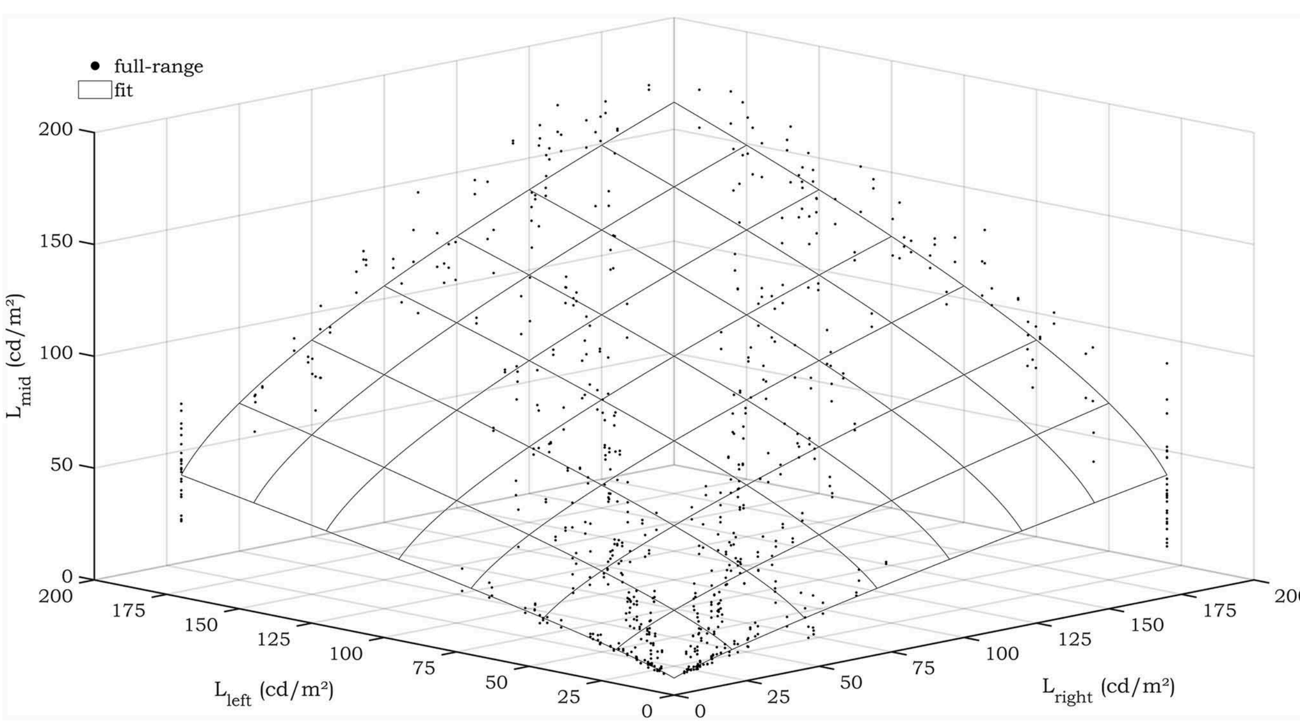

Fig. 8. Adjusted luminance values of the individual observers (dots) for the full-range in the improved PS method. The black grid represents the fitted general model.

3.2.3. Brightness perception: General model

The PS general model (4) was fitted to all individual observer data obtained using the improved PS method and is shown in Fig. 8 as a black grid. The luminance values adjusted by each observer are plotted as black dots. The fit had a very high R^2 of 0.958 with a value and CI for the exponent b of 0.384 and $[0.346, 0.422]$, respectively. Although the b value is slightly smaller compared to the one obtained for the full-range in the standard method (0.403), their confidence intervals overlap, indicating that similar results are obtained with both methods.

Finally, a PS general model was also fitted to the improved PS data, but this time only to the data of the first three subdivision levels, as the 4th level had substantially more average observer bisection errors. The fit had a very high R^2 of 0.942 with a value and CI for the exponent b of 0.380 and $[0.339, 0.421]$, respectively. This indicates that an improved PS run for the full range with three subdivision levels is sufficient to achieve satisfactory results.

3.3. Standard versus improved method

Both methods were compared, but only for the full- & low-range, as previous results showed that observers could not reliably estimate the mid- & high-ranges. The improved PS method reduces cumulative errors, average observer bisection errors of the 3rd subdivision level were less than 0.5%, compared to 3.8% for the standard PS method with the full-range. The improved method also obtained a similar b value fitted to the average observer data compared to the standard method. Boxplots of the b values fitted to the individual observer data are shown in Fig. 9 for the full-range data obtained with both methods and for the low-range obtained with only the standard method. A substantial inter-individual spread of brightness exponents b is visible within each boxplot. The inter-quartile ranges for the three conditions mentioned above are 0.331, 0.199 and 0.282, respectively. Literature shows that the brightness power exponent can vary a lot across individuals. For example, in (Tsubomi et al. 2012), the "inner psychophysics" of brightness perception within individuals was investigated. The power function exponents for subjective brightness ratings varied from 0.14 to 0.46 with a mean of 0.32 for nine observers.

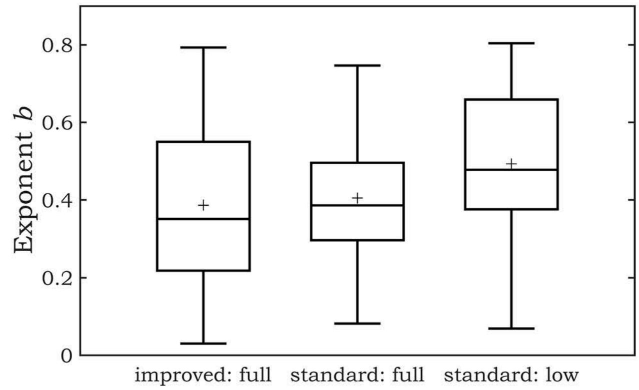

Fig. 9. Boxplots of the exponent b values fitted to the individual observers' data using the general model, applied to the full-range for both methods (improved & standard) and the low-range for the standard method. The mean is shown as a "+" and the median is a horizontal solid black line.

A linear mixed-effects model (LMM) with the individual observers b values as input has been used to analyse potential differences between both methods and between the luminance range adopted in the experiment. Restricted maximum likelihood was used as estimation method for the model. The fixed effects in the model were the range and method type, no interaction term was used. Each observer had an intercept as only random effect, which was assumed to be normally distributed (Shapiro-Wilk's test: $p = 0.339$). Visual inspection of residual plots did not reveal any deviations from homoscedasticity or normality. Significance of fixed effects was tested using the Wald Chisquare test. Neither method type ($\chi^2(1) = 0.067$, $p = 0.797$), nor range ($\chi^2(1) = 2.829, p = 0.093$) was significant, indicating that similar b values were obtained between both methods for the full-range and between the full- & low-range for the standard method. Both standard and improved are viable methods capable of accurately capturing a brightness scale, although the improved method leads to improved observer accuracy. In theory, a perceptual brightness scale of a single observer is obtainable in less than 5 minutes by using only one luminance range and up to three subdivision levels. However, it is advised to conduct several repeats and if possible use the improved method to reduce cumulative errors.

4. Conclusion

Psychophysical data was obtained to derive a perceptual scale for brightness versus luminance using partition scaling. Psychophysical experiments were conducted with simple neutral discs varying in luminance and presented on a dark background and in a dark environment on a calibrated monitor. Four luminance ranges were investigated, a full-range, from 5 cd/m^2 to 175 cd/m^2 and three subranges (low-, mid- & high-range) with luminance intervals equally spaced over the full luminance range.

Average observer brightness perception was modelled for each luminance range using a brightness power law. The full- & low-range had similar brightness exponents, whereas the mid- & high-range had substantially higher offsets and implausible power law exponents compared to values reported in literature. The latter two ranges also had higher average observer bisection errors, confirming observer reports of difficulties with completing the bisection task for these ranges, presumably because the luminance difference (and hence perceived brightness) between two anchor discs was too small, of the order of a just noticeable difference. Despite this, the model fitted to the average observer data had R^2 values of 0.99 and higher. This indicates that R^2 on average observer data is not a good estimate to indicate whether or not observers could actually complete the task correctly.

Pooling all ranges and rescaling the data using the full-range as a common scale showed no indication for a range bias. A general model was defined with only the brightness exponent as parameter and which allows the pooling of the unscaled individual ranges. General model fits to the individual observer data of the individual ranges resulted in similar brightness exponent values and substantially lower R^2 compared to fits to the average observer data. Analysis showed no indication of a position bias or starting luminance bias, nor was a range bias found for the full- & low-ranges.

Partition scaling is capable of capturing a brightness scale without the need of any prior knowledge of the psychophysical brightness function. However, PS is not flawless; it can be prone to cumulative errors, as each subsequent bisection interval is dependent on the previously set interval. Results indicate that the initial adjusted luminance value for a luminance range could predominantly influence the fitting results. Therefore, it is advisable to conduct several repeats of the same luminance range or to use the improved PS presented in this paper. Results obtained with the improved PS showed similar brightness exponent values, while fewer average observer bisection errors occurred, indicating improved observer accuracy and a decrease in possible accumulation errors.

Finally, although PS is rarely used in lighting research experiments, this paper showed that PS is a valid method for brightness perception experiments with the advantage that it is fast and reliable. Future research could focus on the validation of PS, not only for investigating brightness perception for simple stimuli, but also for colour perception or other "visual attributes" in general.

Acknowledgments

The authors want to express their gratitude towards the observers and the two anonymous reviewers for their constructive feedback.

Disclosure statement

The authors have no financial interests to declare.

Funding

This research was supported by Impulse Fund KU Leuven (IMP/16/027 – "Correlaties tussen objectieve en subjective descriptoren van binnenverlichting"), granted to W. R. Ryckaert and Marc Dujardin.

ORCID

Laurens Van de Perre http://orcid.org/0000-0001-5778-3198
Peter Hanselaer http://orcid.org/0000-0002-6060-3706
Kevin A.G. Smet http://orcid.org/0000-0003-3825-6274

References

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using {lme4}. J Stat Softw. 67(1):1–48.
Brainard DH. 1997. The psychophysics toolbox. Spat Vis. 10 (4):433–436.
Garner WR. 1954. A technique and a scale for loudness measurement. J Acoust Soc Am. 26(1):73–88.
Gescheider. 1997. Psychophysics: The fundamentals. 3rd ed. London (UK): Lawrence Erlbaum Associates.

Katoh N, Deguchi T, Berns RS. 2001. An accurate characterization of CRT monitor (I) verifications of past studies and clarifications of gamma. Opt Rev. 8(5):305–314.

Kingdom FAA, Prins N. 2016. Psychophysics. A practical introduction. 2nd ed. Academic Press.

Laming DRJ, Donald RJ. 1997. *The measurement of sensation*. Oxford (UK): Oxford University Press.

Poulton EC. 1979. Models for biases in judging sensory magnitude. Psychol Bull. 86(4):777–803.

R Core Team. 2019. R: A language and environment for statistical computing. Vienna (Austria).

Stevens SS. 1957. On the psychophysical law. Psychol Rev. 64 (3):153–181.

Stevens SS. 1961. To honor fechner and repeal his law. Science. 133(3446):80–86.

Stevens SS. 1975. Psychophysics: Introduction to its perceptual, neural, and social prospects. New York (NY): Wiley.

Stevens SS, Stevens JC. 1960. The dynamics of visual brightness. Psychophysical Laboratory Monograph. Boston (MA): Harvard University.

Stevens SS, Volkmann J. 1940. The relation of pitch to frequency: A revised scale. Am J Psychol. 53(3):329.

Teghtsoonian R, Teghtsoonian M. 1978. Range and regression effects in magnitude scaling. Percept Psychophys. 24 (4):305–314.

Tsubomi H, Ikeda T, Osaka N. 2012. Primary visual cortex scales individual's perceived brightness with power function: Inner psychophysics with fMRI. J Exp Psychol. 38(6):1341–1347.

Whittle P. 1992. Brightness, discriminability and the 'Crispening Effect'. Vision Res. 32(8):1493–1507.

Withouck M, Smet KAG, Hanselaer P. 2015. Brightness prediction of different sized unrelated self-luminous stimuli. Opt Express. 23(10):13455–13466.

Appendix: Observer instructions

The following sections show the instructions that observers received for each PS method. The observer instructions were given verbally either in Dutch or English depending on the observer's preference. Text highlighted in italic were the verbally given instructions to the observer.

PS

"*You will be presented with three (or two) luminous circles. The left circle will be somewhat dimmer than the right circle* (this is reversed in session two). *The middle circle* (adjustable disc) *will either be very bright or dark (invisible). You can control the brightness intensity of the middle circle by using the up or down arrow key, this will increase or decrease the brightness, respectively. Pressing down the control-key along with an arrow key will increase the transition speed of the middle circle going darker or brighter depending on which arrow key is pressed. You can try this out now*." After confirmation that the observer understood the controls, the observers' task was explained: "*The task is to adjust the middle circle so that the brightness step between the left and middle circle and the middle and right circle is equivalent, thus essentially centring the brightness perception of the middle circle between the left and right circle as good as you can. When you are satisfied with the result you can press the space key to continue to the next task. There is no time limit imposed for adjusting the middle circle. It is possible to take a break or return to the previous task by informing me* (the experimenter). *Now we will start with an introduction if you have no further questions*." After the introduction was completed, the observer was asked if everything is clear. "*Now we will begin the actual experiment, please note that in the beginning it will be easy to see the brightness differences between the left and right circle, as the experiment progresses this may become more difficult, try to do your best at all times and when you can't visibly see any difference try to place the brightness in the middle and proceed to the next task.*".

Improved PS

The observer instructions for the improved PS method were similar to those of the standard PS method, except that the left circle could be either darker or brighter than the right circle. Therefore, the only difference was the following: "*You will be presented with three (or two) luminous circles. The left circle can be somewhat dimmer than the right circle or visa-versa.*"