Lighting Res. Technol. 2015; **0**: 1–19

Memory and preferred colours and the colour rendition of white light sources

KAG Smet PhD and **P Hanselaer** PhD
Light and Lighting Laboratory, KU Leuven, Ghent, Belgium

Received 13 November 2014; Revised 19 December 2014; Accepted 26 December 2014

The application of memory and preferred colours to colour rendition evaluation of white light sources is reviewed. Four metrics are discussed: Sanders' preferred colour index, Judd's flattery index, Thornton's colour preference index and Smet's memory colour rendition index. Following a review of the metrics themselves, the paper continues with a discussion of their predictive performance in terms of agreement with psychophysical data on visual appreciation and naturalness perception. Their performance was also compared to that of several other colour rendition metrics and the impact on the predictive performance of a metric's emphasis on chroma enhancement has been evaluated.

1. Introduction

Memory colours refer to the colours associated with familiar objects in long-term memory, e.g. the yellow of a ripe banana, and preferred colours are the colours one would like (prefer) objects to have. Memory colours were first introduced by Hering in the late 19th century who stated that we view the world through the spectacles of memory.[1] Since then, memory colours and preferred colours have been studied by many researchers.[2–25]

Because of their potential use as an internal reference, they have especially been of interest to areas of colour research that involve assessment of object colour appearance, colour quality and colour reproduction[3,5,9,15,20,25–28] and colour rendition of light sources.[21,29–33]

This paper focuses on the area of colour rendition evaluation based on memory and preferred colours. More specifically, it shortly reviews, in chronological order, Sanders' preferred colour index R_p,[30] Judd's flattery index R_f,[29] Thornton's colour preference index (CPI)[33] and Smet's memory colour rendition index R_m.[21,32] Finally, the ability of these and several other metrics to predict subjective aspects of colour rendition, such as visual appreciation and naturalness, is discussed based on visual data obtained from 17 psychophysical studies.[31,34–49]

For an overview of past and recent work on colour rendition in general and a review of some of the other colour rendition metrics, the reader is referred to the literature.[50–54] A review of the work on memory and preferred colours in general can be found in Smet *et al.*[55]

2. Memory and preferred colour rendition metrics

Memory and preferred colour rendition metrics are based on the assumption that the colour rendition or colour quality of a light source improves when the colour of familiar objects is rendered more closely to

Address for correspondence: Kevin Smet, Light and Lighting Laboratory, KU Leuven @ campus Ghent, Gebroeders Desmetstraat 1, B9000 Ghent, Belgium.
E-mail: Kevin.Smet@kuleuven.be

10.1177/1477153514568584

what is expected or preferred. The following subsections review the four only existing colour rendition metrics based on either memory or preferred colours.

2.1. Sanders' preferred colour index, R_p

In 1959, Sanders developed a colour rendition metric based on preferred colours[30] and colour tolerance ellipses obtained in a series of psychophysical rating experiments.[18] In the study, six familiar objects – 'hand', 'face', 'tea', 'butter', 'potato chips' and 'beefsteak' – were investigated under adapting illumination with chromaticities approximately corresponding to Commission Internationale de l'Eclairage (CIE) illuminants B and C. The objects were presented to a group of observers in at least 20 different colours, by illuminating them with light composed of different ratios of blue, green and pink light obtained from three pairs of fluorescent light sources. The observers were asked to rate the apparent colour on a five-point scale: 'good', 'fair to good', 'fair', 'fair to unsatisfactory' and 'unsatisfactory'. For calculation purposes, values of, respectively, 100, 75, 50, 25 and 0 have been assigned to these ratings. For illustrative purposes, the preferred colours and colour tolerance ellipses obtained under illuminant B are shown in Figure 1(a), and similar graphs for illuminant *C* can be found in Sanders.[18]

Sanders evaluated colour rendition by comparing the chromaticities of the six familiar objects illuminated by the test source and corrected by a Judd-type (translational) chromatic adaptation transform with their preferred chromaticities.

To reduce the error of the chromatic adaptation, he selected either the illuminant B or C preferred colour data sets, whichever the test lamp chromaticity was closest to. Practically, the comparison was made by first calculating the subjective colour deviations $\Delta S/r_\theta$, with ΔS the chromaticity difference between the apparent object chromaticity (x,y) and the preferred chromaticity (x_e,y_e)

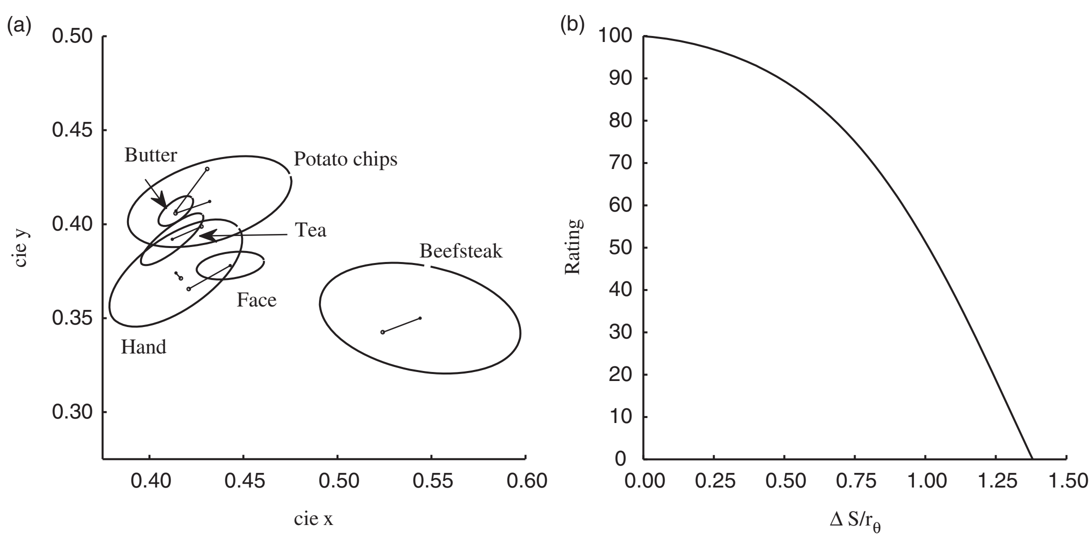

Figure 1 (a) Preferred colours and colour tolerance ellipses under illuminant B of the six familiar objects investigated by Sanders. (b) Relationship between the colour deviation and experimentally determined colour rendition rating (reproduced from Sanders[18])

and with r_θ the radius of the tolerance ellipses in the same direction

$$\Delta S/r_\theta = g_{11}(x - x_e)^2 + g_{22}(y - y_e)^2 + 2g_{12}(x - x_e)(y - y_e) \quad (1)$$

The parameters g_{ij} (published in Sanders[30]) determine the size, shape and orientation of the experimentally determined tolerance ellipses.

Finally, as the unsuitability of an object colour is not necessarily proportional to the subjective colour deviation, Sanders rescaled it using the empirically derived relationship between the subjective colour deviations and the colour rendition ratings obtained in his experiments (see Figure 1(b)). However, to assign a value of 85 to illuminant B, $\Delta S/r_\theta$ has to be divided by 2.06 prior to rescaling. A general colour rendition index is then defined as the average of the results of the six familiar objects. Unfortunately, the latter only span the yellow-red region of colour space which may, as acknowledged by Sanders, result in high scores for light sources that render blues, greens and magentas poorly (and vice versa). More details can be found in Sanders' 1959 papers.[18,30]

2.2. Judd's flattery index, R_f

In 1967, Judd proposed his flattery index, R_f, as a supplement to the CIE colour rendering index,[56,57] because of concerns that the CIE "*color rendering index of a light source may correlate poorly with public preference of the source for general lighting purposes*".[29]

Although Judd's flattery index is based on memory and preferred colours, it does not use actual memory and preferred chromaticity of familiar objects as reference. Instead, it closely follows the calculation scheme of the CIE colour rendering index, but corrects the chromaticities of a number of Munsell samples (CIE test samples 1–8, 13 and 14) illuminated by the reference illuminant with a preferred chromaticity shift. The latter were obtained from the memory and preferred colour data obtained by Bartleson,[2] Sanders[18] and Newhall *et al*.[58] Judd states that although the preferred shifts "*refer to natural overcast sky light and to artificial illuminants (such as CIE source C) intended to approximate it*",[29] they could also be used for light sources with correlated colour temperatures of 3500 K and higher. For sources with lower correlated colour temperatures (CCTs), suitable adjustments would probably be required.[29] The preferred chromaticity shifts are illustrated in Figure 2. As can be seen, the preferred chroma shifts are generally in the direction of increased saturation.

It may be noted that the calculation scheme of the CIE colour rendering index as proposed in the 1960s[56,57] is slightly different from the current CIE colour rendering index standardized in 1974.[59] One of the differences is that the latter incorporates a von Kries-type chromatic adaptation transform, while Judd's flattery index, as proposed in his 1967 paper, uses a translational or Judd-type chromatic adaptation transform to account for any difference in chromaticity between the test source and its reference illuminant.

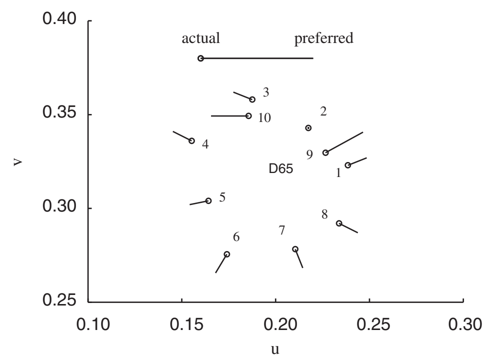

Figure 2 Judd's preferred chromaticity shift (full magnitude) for the Munsell samples # 1–8, 13 and 14 (reproduced from Judd[29])

Judd's flattery index, R_f, is calculated as

$$R_f = 100 - 4.6\overline{\Delta E_{f,k}} \quad (2)$$

with $\overline{\Delta E_{f,k}}$ the weighted arithmetic mean of the chromaticity differences between the chromaticity of the 10 Munsell samples illuminated by the test source and their chromaticity under the reference illuminant, corrected by one-fifth of the preferred chromaticity shift. Judd kept the 4.6 scaling factor of the CIE colour rendering index but used only one-fifth of the calculated preferred chromaticity shift to assign a value of 90 to the reference illuminant. Light sources can therefore score higher than their reference illuminants, i.e. reference illuminants are not optimal, but not higher than 100. Finally, chromaticity is calculated in the CIE 1960 uniform colour space. For further details, the reader is referred to Judd's 1967 paper.[29]

2.3. Thornton's preference index, *CPI*

With the exception of a few differences, the CPI[33] proposed by Thornton in 1974 is very similar to Judd's flattery index. Like Judd's index, it does not use actual memory or preferred colours, but instead corrects the chromaticity of a number of Munsell samples illuminated by the reference illuminant. Thornton's CPI uses only the first eight Munsell samples and also keeps the original magnitude of the preferred chromaticity shift calculated by Judd.[29] In addition, Thornton weighted all samples equally. Finally, the maximum preference score of a light source is 156 and illuminant D65 was assigned a value of 100. Thornton's CPI is calculated as follows

$$CPI = 156 - 7.18\overline{\Delta E} \quad (3)$$

where $\overline{\Delta E}$ is the arithmetic mean of the colour shift in the CIE 1960 uniform colour space.

2.4. Smet's memory colour rendition index, R_m or MCRI

The memory colour rendition index, R_m (also known as MCRI) is a memory colour metric that assesses a light source's colour rendition by comparing the rendered colours of a number of familiar objects to their actual memory colours using empirically derived similarity functions. The latter describe the psychophysical response to a chromaticity deviation from the memory colour, implicitly taking potential differences in chroma and hue tolerance into account.

Memory colours and associated similarity functions of 10 familiar objects with good hue coverage ('green apple', 'ripe banana', 'orange', 'dried lavender', 'smurf®', 'strawberry yoghurt', 'sliced cucumber', 'cauliflower', 'Caucasian skin' and 'neutral gray') were derived from observer ratings obtained in visual experiments.[60] Each familiar object was presented in over 100 different chromaticities (with approximately constant luminance) to a panel of 32 colour normal observers. The observers had to rate the apparent object colour, on a five-point scale ('1: very bad', '2: bad', '3: neutral', '4: good', '5: very good'), with reference to what they imagine the object looks like in reality. The apparent colour was changed by altering the luminous flux of the R(ed) G(reen) B(lue) A(mber) light-emitting-diodes (LED) illuminating them while hiding all clues to the colour of the illumination. The observer adaptation state was kept approximately constant by presenting the object in front of a self-luminous back panel with a CCT of 5600 K.

For each object a similarity function was obtained by normalizing the bivariate Gaussian model fitted to the observer ratings.[21,32] For illustrative purposes the Gaussian rating model for a 'green apple' is shown in Figure 3(a). Cross-sections of the similarity functions corresponding to a unit Mahalanobis distance (1d-contours) are illustrated in Figure 3(b).

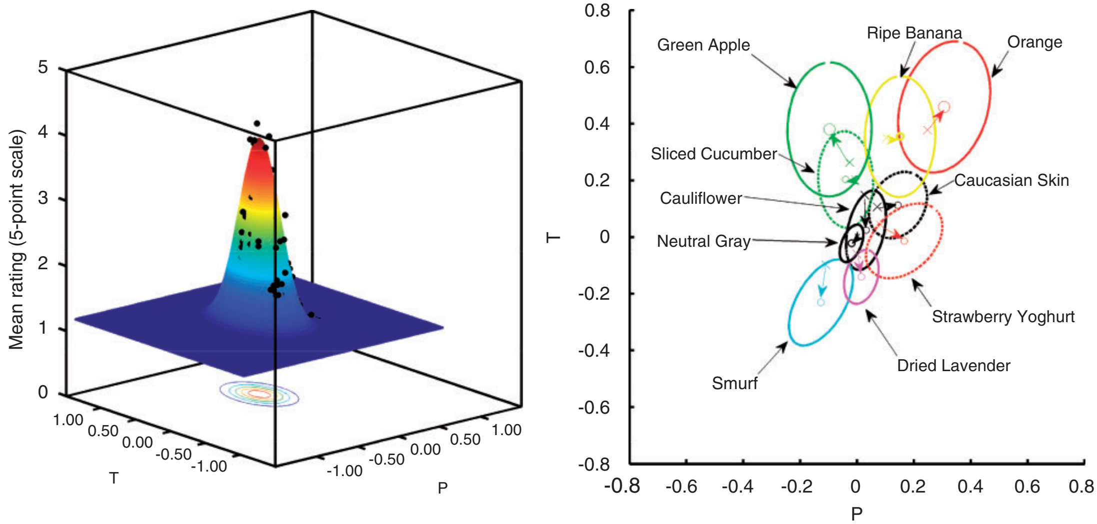

Figure 3 (a) The Gaussian rating model and mean observer ratings for an apple. (b) The 1d-elliptical contours of the 10 similarity functions. Centres are represented by (+), the typical object chromaticity under D65 by (x)

For comparison, the typical object chromaticity under D65 is also plotted.

Two important conclusions can be drawn from Figure 3. One, the major axis of the similarity functions is approximately orientated along the chroma direction suggesting a higher tolerance for deviations in chroma than for hue. And two, compared to the objects' typical colour under daylight, memory colours of most familiar objects tend to be more saturated[21] consistent with reports in the literature.[2,20,58,61] It might also explain why light sources that increase object gamut or saturation are, up to a point, typically perceived to have better subjective colour rendition quality (e.g. 'preference').

The memory colour rendition metric, R_m is calculated as follows:

1) For each familiar object illuminated by the test source, the 10° corresponding chromaticity under D65 is calculated in IPT[62] colour space ($P_i T_i$). P and T are, respectively, the red-green and yellow-blue axis of the IPT colour space. Chromatic adaptation is taken care of by the CAT02 transform as defined in CIECAM02, whereby the degree of adaptation D is determined by the luminance of the adaptation field. If unknown, a value of 0.90 is recommended, as chromatic adaptation is rarely complete, especially for light sources having CCTs much different from D65.[63–65]
2) The specific degree of similarity S_i with the memory colour of a familiar object is determined using the similarity function $S_i(P,T)$

$$S_i(P_i, T_i) = e^{-\frac{1}{2}\left[a_{i,3}\left(P_i-a_{i,1}\right)^2+2a_{i,5}\left(P_i-a_{i,1}\right)\left(T_i-a_{i,2}\right)+a_{i,4}\left(T_{i,1}-a_{i,2}\right)^2\right]} \times (i=1\ldots10) \qquad (4)$$

with a_1–a_5 fitting parameters describing the function's centroid, shape, size and orientation. Note that the bracketed part is a non-Euclidean distance measure – called a Mahalanobis distance – that implicitly takes differences in chroma and hue tolerances into account.

3) The general degree of similarity S_a is obtained as the geometric mean of the S_i values.
4) The 0–1 range of the general S_a is rescaled to a more familiar 0–100 range using a sigmoid function

$$R_m = 100 \cdot \left(\frac{2}{e^{p_1 |\ln(S_a)|^{p_2}+1}} \right)^{p_3} \qquad (5)$$

with the rescaling parameters p_1–p_3 such that the CIE illuminants F4 and D65 have, respectively, R_m values of 50 and 90 and such that $S_a < 0.5$ corresponds to $R_m \approx 0$.

For a step-by-step overview with all required equations and parameter values, the reader is referred to Smet *et al.*[66]

3. Performance of colour rendition metrics

The predictive performance of the memory colour rendition metrics discussed earlier was investigated. In addition to the memory colour-based metrics, several other colour rendition metrics have been included as well for comparison. These included both metrics specifically designed to predict more subjective aspects of colour rendition (naturalness, preference, vividness, etc.), as well as so-called fidelity-type metrics that attempt to provide an objective measure of colour rendition with respect to a broadband reference illuminant. Note that the latter need not be optimal in terms of any of the subjective aspects. Fidelity metrics are however included to illustrate that they are indeed unsuitable to predict colour rendition aspects such as preference and naturalness.

The following colour rendition metrics have been investigated:

(a) Preferred colour based: Sanders' Preferred Colour Index R_p,[30] Judd's Flattery Index R_f[29] and Thornton's CPI.[33]
(b) Memory colour based: Smet's Memory Colour Rendition Index R_m or (MRCI).[21,32]
(c) Colour fidelity based: the CIE R_a,[59] *CRI2012* $R_{a,2012}$[67] and CQS Q_f (*v9.0*)[68] (CQS *v9.0* is unpublished, but is almost identical to *v7.5* published in Davis and Ohno.[68] Calculations were performed following the procedure outlined in an Excel CQSv9.0 calculator distributed by Dr Y. Ohno from the US National Institute of Standard and Technology).
(d) Chroma enhancement/gamut-expansion based: CQS (*v9.0*) Q_p and CQS (v9.0) Q_g,[68] the feeling of contrast colour rendering Index (*FCI*)[69] and the Gamut Area Index (*GAI*).[70]
(e) Other (e.g. mixed fidelity and chroma or gamut based): CQS Q_a (*v9.0*),[68] and the $GAIR_a$, the geometric mean of the *CIE* R_a and *GAI*.[32]

Performance of the colour rendition metrics was assessed by a meta-analysis of the Spearman correlation coefficients between the metric predictions and the psychophysical ratings of colour rendition obtained from several studies in literature.[31,34–49] A Spearman rank correlation coefficient was used as it does not make any assumptions about the data. In addition, it provides a measure of the ability of a metric to correctly rank light sources. A metric with high Pearson (but low Spearman) correlation would be of no practical use.[31] The 17 studies have investigated one or more subjective aspects of colour rendition. Twenty one experiments were devoted to the study of visual appreciation (preference and attractiveness)[31,34–36,39–49] and 15 to naturalness.[31,34,35,37–42,44,45,49] For each aspect a meta-analysis was conducted.

As six visual experiments used object sets that lacked full hue coverage (mostly blue to magenta),[34,35,41,48,49] possible bias due to mismatch between the experiment object set and a metric sample set[30,31] was minimized by limiting the hue range of the latter to that of the former prior to calculating the metric scores. A similar approach was also used in Smet *et al*.[32,50] Finally, it may be noted that the light sources in the psychophysical studies had primarily CCTs of approximately 3000 K, as it is generally much easier to find an appropriate smooth broadband reference light source (incandescent or halogen).

3.1. Statistical analyses

3.1.1. Meta-analysis and corrections for error and artefacts

A meta-analysis is a statistical method to estimate the true strength of association – in this case the Spearman correlation – between variables by combining data from several studies and by, if possible, correcting for sampling error and study artefacts. The meta-analysis followed the method of Hunter–Schmidt (HS), whereby the true correlation is estimated by a weighted average correlation.[71]

In a 'bare-bones' meta-analysis only the sampling error or within-study variance is corrected for and the weights are the number of samples (N) in each study. However, other types of error and study imperfections (called artefacts) should be corrected for when possible, as they have a tendency to attenuate the true correlation and typically also lead to an underestimation of the true variance.

Note that, as stated by Hunter and Schmidt,[71] "*the artifact attenuation is caused by real imperfections in the study design. The attenuation of the true correlation will thus occur whether we can correct for it or not*."

In the present analysis, the following errors and artefacts were corrected for:

1) Sampling error (within-study variance)
2) Study heterogeneity (between-study variance): Accounted for by adjusting the study weights from N to the optimal weights $1/(\tau + N^{-1})$, with τ the HS heterogeneity estimator.[72]
3) Range restriction/enhancement: When one of the variables over which the correlation is calculated has only values within a limited range compared to the overall or true range, the correlation tends to be reduced. Hunter and Schmidt state that "*correlations are directly comparable across studies only if they are computed on samples from populations with the same standard deviation on the independent variable*".[71] Therefore, range correction has been applied with respect to the range observed over all experiments. For each study, the range restriction parameter ux has been estimated as the ratio of the standard deviation of the metric scores for that study to that of the pooled metric scores of all studies available.
4) Attenuation due to inter-rater idiosyncrasy (halo): Inter-rater variability was treated as measurement error and estimated as (*1 – inter-rater STRESS/100*), when available. When missing, inter-rater variability was estimated as the mean of the studies that did have that data. The mean inter-rater STRESS across all studies was 25, comparable to typical values obtained in colour discrimination studies.[73]
5) Sample correlation bias: The sample correlation r is a statistically biased estimator of the population correlation ρ and was corrected by the linear ($r < 0.7$) or non-linear ($r \geq 0.7$) attenuation factor defined in equations (3.26) and (3.27) in Hunter and Schmidt.[71]

For an in-depth discussion on these corrections and on meta-analysis in general, the reader is referred to Hunter and Schmidt.[71]

3.1.2. Significance testing of metric cross-comparisons

All metric performances were also cross-compared. Statistical differences were determined using the confidence interval method of Zou[74] for dependent overlapping correlations under the null hypothesis of equal corrected average correlations, H_0: $\overline{r_c i} - \overline{r_c j} = 0$

(whereby i and j run over all metrics). A typical significance level of $\alpha = 0.05$ was selected *a priori*. The method tests whether the $100 \times (1\text{–}\alpha)\%$-confidence interval (CI) contains zero, in which case H_0 cannot be rejected.

3.1.3. Multidimensional scaling (MDS) analysis

To verify the results of the correlation meta-analysis a MDS-based analysis was conducted on the metric scores and the visual ratings for each aspect of colour rendition. The MDS attempts to find relationships in a set of data such that the Euclidean distances between the objects in the data set are preserved when represented in a lower dimensional space. A similar analysis has been done by Houser *et al*.[51] on a set of metrics scores of a large number of light source spectra. However, they only took metric scores into account, but no visual data.

In the present analysis, the MDS minimized the stress of the distance matrix containing the Euclidean distances between all cross-compared scores (metric and visual rating). Matrix stress is an indicator of how poorly the distances between scores are preserved when mapping to a lower dimensional space. Values smaller than 0.2 are considered a good mapping.[51]

In addition, the coefficient of variation (R^2), which quantifies the amount of variance explained by the MDS mapping was also calculated, as well as the relative contribution of each MDS axis.

As the various metrics and ratings used different scales, the metric scores and the visual ratings for each study were transformed to z-scores before applying the MDS.[51]

After the MDS, the Euclidean distances between the lower dimensional representation of the metric scores and that of the visual rating were determined. As smaller distances indicate better agreement with the visual rating, they can be used to verify and complement the results of the correlation analysis. The strength of the agreement between the MDS derived metric-rating distances and the metric correlation was quantified using the Spearman correlation coefficient. Good agreement testifies to the reliability of the estimated metric performance and the conclusions drawn.

3.2. Results

3.2.1. Correlation meta-analysis and metric performance comparisons

The weighted average artefact-corrected Spearman correlation coefficients for the different metrics with the visual appreciation and naturalness ratings are presented in Table 1. For comparison, the results of the bare-bones analysis and the intermediate step-by-step corrected average correlations are also shown. For a quick and easy overview, the final results obtained by applying all corrections are also illustrated in Figure 4.

From Table 1, it is clear that the artefact corrections had a de-attenuating effect for almost all metrics, with the exception for the correlation between the FCI and naturalness (which was attenuated to correct for the inflation due to range enhancement).

It is also clear (see also Figure 4) that the metrics vary substantially in their predictive performance. The statistical significance of these differences in performance was determined in a series of cross-comparisons. The CI-bound closest to zero, henceforth referred to as CI_0, of the CI tests on the correlation difference (H_0: $\overline{r_c i} - \overline{r_c j} = 0$) are summarized in Table 2. When there is a statistically significant difference, the CI_0 has been underlined.

3.2.2. MDS analysis

Finally, the results of the correlation meta-analysis were verified using an analysis based on a MDS of the metric scores and visual ratings (see Figure 5). The mapping to a two-dimensional space of the visual appreciation and naturalness data resulted in excellent distance matrix stress values of, respectively,

Table 1 Metric performance: The weighted average (artefact-corrected) Spearman correlation coefficients $\overline{r_c}$ for visual appreciation and naturalness. The intermediate results when sequentially applying the artefact corrections are also shown (bottom to top)

Applied corrections	Visual appreciation												
	R_a	$R_{a,2012}$	Q_f	Q_a	Q_p	Q_g	*FCI*	*GAI*	$GAIR_a$	R_p	R_f	*CPI*	R_m
All	0.09	0.41	0.16	0.38	0.81	0.76	0.67	0.74	0.77	−0.26	0.57	0.77	1.00
2,3,4	0.06	0.34	0.12	0.32	0.78	0.71	0.61	0.68	0.74	−0.27	0.51	0.73	1.00
2,3	0.06	0.31	0.12	0.29	0.67	0.62	0.54	0.63	0.65	−0.23	0.44	0.63	0.90
2	0.02	0.30	0.11	0.29	0.67	0.62	0.53	0.50	0.47	−0.24	0.47	0.61	0.79
Bare-bones	0.02	0.33	0.13	0.31	0.67	0.62	0.56	0.51	0.47	−0.32	0.48	0.61	0.79
Applied corrections	Naturalness												
	R_a	$R_{a,2012}$	Q_f	Q_a	Q_p	Q_g	*FCI*	*GAI*	$GAIR_a$	R_p	R_f	*CPI*	R_m
All	0.60	0.61	0.63	0.70	0.75	0.30	0.13	0.39	0.94	0.16	0.72	0.70	0.71
2,3,4	0.52	0.55	0.56	0.62	0.69	0.24	0.08	0.31	0.89	0.11	0.64	0.65	0.65
2,3	0.44	0.46	0.47	0.52	0.61	0.21	0.07	0.28	0.80	0.14	0.55	0.56	0.57
2	0.39	0.43	0.40	0.50	0.60	0.26	0.11	0.23	0.62	0.07	0.55	0.55	0.49
Bare-bones	0.36	0.43	0.38	0.50	0.60	0.30	0.20	0.26	0.61	−0.04	0.55	0.55	0.49

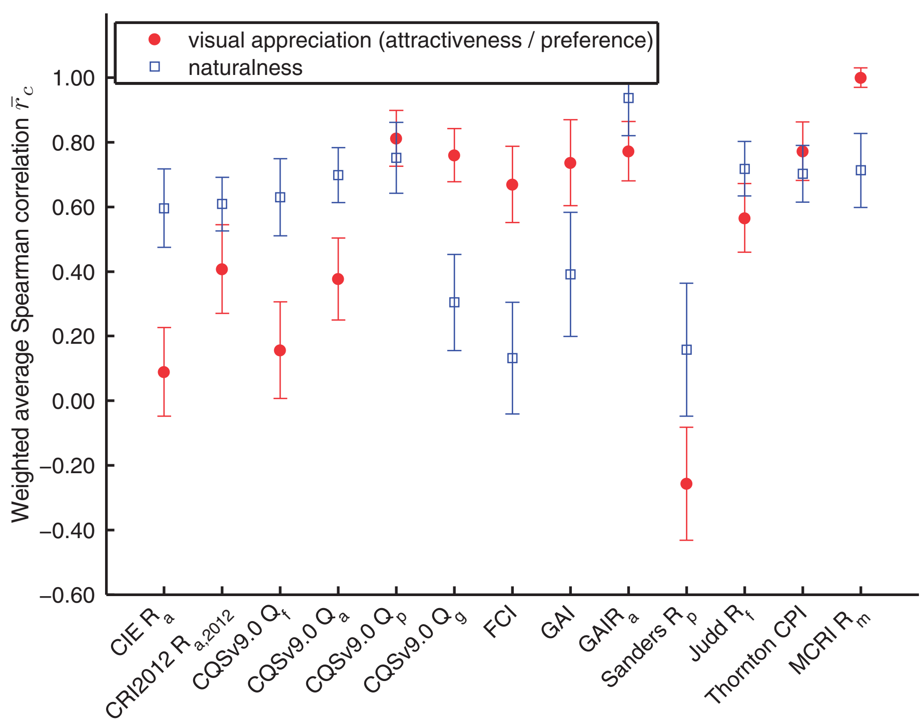

Figure 4 Metric performance: The weighted average artefact-corrected Spearman correlation (with standard error) between the metric predictions and the observer ratings for visual appreciation (red filled circles) and naturalness (blue open squares) obtained in 21 and 15 experiments, respectively

Table 2 Results for the cross-comparisons of metric performance (H_0: $\overline{r_c i} - \overline{r_c j} = 0$). The values are the confidence interval bound closest to zero. Underlined values signify statistically significant differences, i.e. zero was within the confidence interval bounds. Results for visual appreciation and naturalness are, respectively, shown in the upper and lower triangles of the table

	R_a	$R_{a,2012}$	Q_f	Q_a	Q_p	Q_g	*FCI*	*GAI*	$GAIR_a$	R_p	R_f	*CPI*	R_m
R_a		<u>−0.19</u>	0.01	<u>−0.10</u>	<u>−0.45</u>	<u>−0.31</u>	<u>−0.13</u>	<u>−0.20</u>	<u>−0.42</u>	<u>0.01</u>	<u>−0.23</u>	<u>−0.33</u>	<u>−0.68</u>
$R_{a,2012}$	−0.12		<u>0.14</u>	−0.13	<u>−0.15</u>	<u>−0.02</u>	0.13	0.10	<u>−0.07</u>	<u>0.24</u>	0.05	<u>−0.03</u>	<u>−0.34</u>
Q_f	−0.08	−0.13		<u>−0.05</u>	<u>−0.36</u>	<u>−0.22</u>	<u>−0.06</u>	<u>−0.11</u>	<u>−0.32</u>	<u>0.05</u>	<u>−0.16</u>	<u>−0.25</u>	<u>−0.58</u>
Q_a	−0.06	0.00	−0.08		<u>−0.21</u>	<u>−0.07</u>	0.09	0.05	<u>−0.15</u>	<u>0.21</u>	<u>−0.04</u>	<u>−0.10</u>	<u>−0.39</u>
Q_p	−0.27	−0.21	−0.31	−0.33		−0.13	−0.13	−0.24	−0.15	<u>0.67</u>	<u>0.11</u>	−0.10	<u>−0.04</u>
Q_g	0.17	0.06	0.14	<u>−0.01</u>	<u>−0.18</u>		−0.06	−0.17	0.21	<u>0.57</u>	−0.04	0.10	<u>−0.08</u>
FCI	0.04	<u>−0.07</u>	<u>0.00</u>	<u>−0.17</u>	<u>−0.26</u>	<u>−0.01</u>		0.14	0.22	<u>0.39</u>	−0.20	0.14	<u>−0.09</u>
GAI	0.32	0.23	0.29	0.16	<u>−0.02</u>	−0.11	<u>0.01</u>		0.22	<u>0.49</u>	−0.18	0.21	<u>−0.01</u>
$GAIR_a$	−0.05	<u>0.01</u>	−0.09	−0.13	<u>0.13</u>	<u>0.21</u>	<u>0.36</u>	<u>0.13</u>		<u>0.69</u>	<u>0.01</u>	−0.21	<u>−0.07</u>
R_p	0.04	0.03	0.00	<u>−0.05</u>	<u>−0.21</u>	0.44	−0.62	0.38	<u>−0.47</u>		<u>−0.41</u>	<u>−0.61</u>	<u>−0.94</u>
R_f	−0.17	−0.13	−0.20	−0.20	0.24	<u>0.14</u>	<u>0.28</u>	−0.04	0.06	<u>0.11</u>		0.01	<u>−0.24</u>
CPI	−0.26	−0.20	−0.28	−0.30	<u>−0.05</u>	<u>0.17</u>	<u>0.28</u>	<u>0.01</u>	<u>−0.04</u>	<u>0.12</u>	0.16		<u>−0.06</u>
R_m	−0.28	−0.21	−0.30	−0.32	0.09	<u>0.19</u>	<u>0.30</u>	<u>0.02</u>	<u>−0.07</u>	<u>0.11</u>	0.21	−0.13	

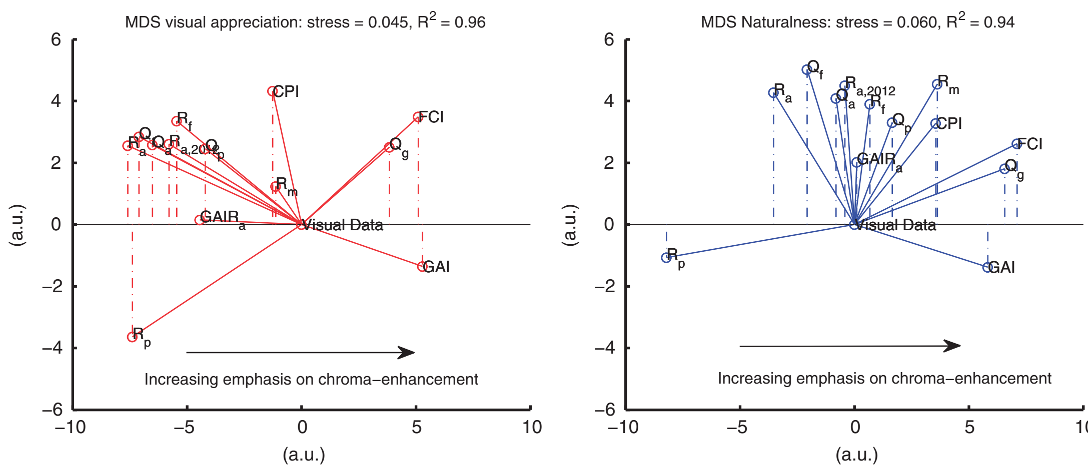

Figure 5 MDS mapping of metric scores to a two-dimensional space centred on the visual rating. Dotted vertical lines were plotted for easy visual metric comparison along the MDS' first dimension. Left: Visual appreciation. Right: Naturalness

0.045 and 0.060, and in excellent R^2 values of 0.96 and 0.94, respectively.

In contrast to Houser *et al.*[51] who interpreted the metric clustering in their MDS analysis as falling along a preference and discrimination or gamut axis, no such two dimensional preference–discrimination structure was readily identifiable.

However, the first MDS axis already accounted for, respectively, 82% and 80% of the total variance of the MDS models for visual appreciation and naturalness and was approximately directed towards increasing emphasis on chroma enhancement in the metric calculation, as becomes clear from Figure 5. Fidelity-type metrics are located

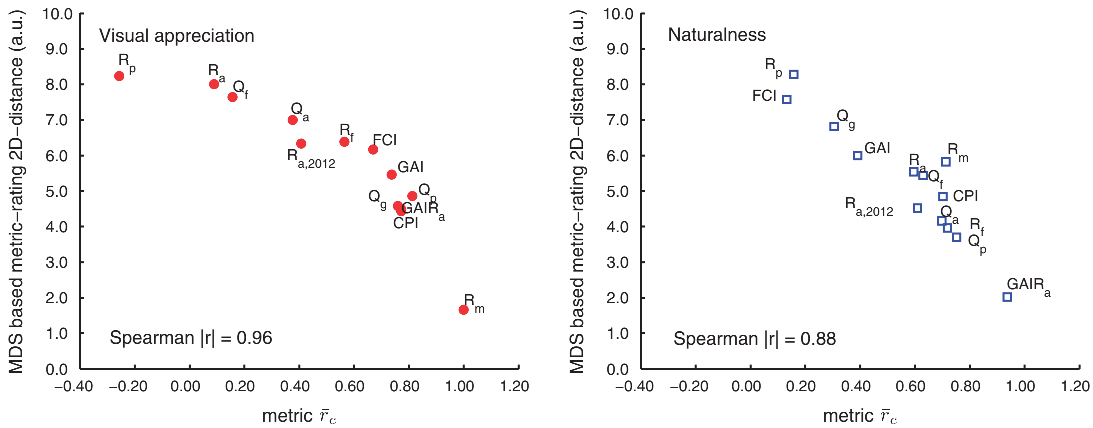

Figure 6 The agreement between the results of the MDS based metric-rating distance and the metric correlation coefficients obtained in the meta-analysis. Left: Visual appreciation. Right: Naturalness

towards the left, gamut area metrics to the right and the others (with saturated reference chromaticities) in between. Furthermore, the direction and interpretation of the first axis is also clearly illustrated by the rank order of the CQS Q_f, Q_a, Q_p and Q_g metrics along this axis.

The relative positions along the first MDS dimension of the naturalness and visual appreciation data also indicate that accurate predictions of the former would require less emphasis on chroma enhancement in the metric calculation than the latter. As shown and discussed more in-depth in the next section, this was also confirmed by the results of the correlation meta-analysis.

With regard to the interpretation of the MDS axes, also note that the preference–discrimination MDS model of Houser *et al.*[51] can be easily reinterpreted to be consistent with the one above by a clockwise rotation of their MDS diagram (see Figure 1 in Houser *et al.*[51]) of approximately 60°. As a bonus, this would also make it consistent with their interpretation of a second MDS analysis on CCT-renormalized metrics, whereby the second axis did not refer to 'preference' (see Figure 4 in Houser *et al.*[51]).

Getting back to the full two-dimensional MDS model, it was found that metric clustering is largely in agreement with the results of the meta-analysis. For example, Q_a and $R_{a,2012}$ are close together, the same for R_a and Q_f. In addition, the former have a slightly smaller metric-rating distance than the latter. All of which is in qualitative agreement with the metric correlation coefficients as can be observed from Figure 4 and Table 1. However, there are other metrics that do not group together but which do show metric-rating distances in qualitative agreement with the magnitude of the correlation coefficients of those metrics (e.g. *FCI*, Q_g, Q_p and *GAI*).

Ignoring structure issues, and focusing only on the two-dimensional, metric-rating distance, a comparison with the metric correlation coefficients obtained in the meta-analysis showed a good overall agreement between the two analysis approaches (see Figure 6).

The magnitudes of the Spearman correlation coefficients between the two approaches to analysis were 0.96 and 0.88 for visual appreciation and naturalness, respectively. The small discrepancy between the two methods for a few metrics is either due to the MDS analysis being directly based on the raw (unweighted and uncorrected) index values and observer ratings, which makes it a rougher and slightly less accurate estimator of the true metric performance; or because the use of ranks in the Spearman correlation ignores the magnitude of the differences between metric and rating values. Whatever the reason, the high correlation between the two methods testifies to the overall reliability of the estimated relative metric performances using either method.

3.3. Discussion

In this section the results presented in the preceding section will be discussed, with the emphasis on the correlation-based performance, because it also provides standard errors on the average performance and hypothesis tests for metric cross-comparisons. Note that the use of the MDS metric-rating distance to estimate performance would have resulted in a qualitatively similar discussion.

As this is a review on memory and preferred colours, the focus will be mainly on the results for the metrics described in Section 2. Each aspect of colour rendition will be discussed in a separate subsection.

3.3.1. Visual appreciation

For visual appreciation, Smet's MCRI R_m had the best performance, according to the MDS analysis as well as the corrected and uncorrected correlation coefficients obtained in the meta-analysis. In fact, after artefact correction, the correlation with visual appreciation was $r_c \pm 1\mathrm{SE}$ (standard error) $= 1.00 \pm 0.03$. A high correlation ($r = 0.88$, uncorrected) had been found before using a more limited set of visual data.[32,50] Applying the same corrections to the correlation of R_m with the limited visual data set gave $r_c \pm 1\mathrm{SE} = 1.00 \pm 0.04$, indicating the old and the new extended data sets on visual appreciation are in good agreement.

The results of the CI test confirmed that the R_m metric was significantly ($\alpha = 0.05$) better than any of the other metrics investigated although it should be noted that, despite a quite substantial nominal difference in correlation values, a significance difference with the *GAI* was only just established. Other metrics with larger, but still rather small CI_0 values were the Q_p and *CPI* metrics.

In contrast to the memory R_m index, the three colour rendition metrics based on preferred colours – Sanders's preferred colour index, Judd's flattery index and Thornton's CPI – showed, respectively, very poor ($r_c \pm 1\mathrm{SE} = –0.26 \pm 0.17$), poor-to-moderate ($r_c \pm 1\mathrm{SE} = 0.57 \pm 0.11$) and moderate-to-good ($r_c \pm 1\mathrm{SE} = 0.77 \pm 0.09$) predictive performance in terms of visual appreciation. Again, the results of the correlation analysis were confirmed by those of the MDS analysis.

Several possible contributors to the lower performance can be identified. First, there is the obvious mismatch between the R_p sample set (red-to-yellow samples) and the experiment object sets. Although, care was taken to avoid this type of bias, it could not be avoided in the case of Sanders' R_p, as this time it was the metric sample set itself that was lacking full hue coverage. Second, the outdated Judd-type chromatic transform also had a major impact on the very poor performance. Updating it to the CAT02 chromatic adaptation transform increased the correlation coefficient of the R_p metric to $r_c \pm 1\mathrm{SE} = -0.16 \pm 0.17$. Other possible contributors to the poor performance are the use of the perceptually non-uniform CIE *xy* chromaticity diagram and the CIE 2° standard observer. The latter is known to be in error in the blue part of the spectrum, which would cause a considerable difference between the instrumental and visual colour matches for many of the LED light sources used in the visual experiments.[75]

In contrast to Sanders' R_p, Judd's flattery index R_f and Thornton's *CPI* have a much better correlation with visual appreciation.

As discussed, both are very similar, with the major difference being the magnitude of the preferred colour shift used to adjust the chromaticity of the test samples under the reference illuminant. Indeed using the full magnitude of the preferred shift in the R_f calculation brings the correlation up to a value of $r_c \pm 1SE = 0.79 \pm 0.07$, which is even slightly better than Thornton's *CPI*. The slight increase compared to the *CPI* is most likely due to the extra two reflectance samples (foliage and skin colour): Setting the weightings in the R_f calculation to 1 and keeping the full preferred colour shift, as is the case in the *CPI* calculation, further increases the correlation to $r_c \pm 1SE = 0.81 \pm 0.08$.

The outdated Judd translational chromatic adaptation transform also had a negative impact on performance. Updating it to the CAT02 transform improved the correlation of the R_f and *CPI* metrics to respectively: $r_c \pm 1SE = 0.59 \pm 0.12$ and $r_c \pm 1SE = 0.82 \pm 0.08$; although the effect is not as large as for Sanders' R_p. This can be understood by considering the magnitude of the adaptive shift required in the metric calculations. In the R_p calculation the reference illuminant is either illuminant B or C, while the R_f and *CPI* metrics use reference illuminants that have the same CCT as the test source. The adaptive shifts typically required, and hence the error made by the outdated Judd translation chromatic adaptation transform, will therefore be much smaller for the latter than the former.

Finally, the lower correlations of the R_f and *CPI* might have in part resulted from ignoring the difference in chroma and hue tolerances by using a Euclidean colour difference equation to estimate the difference in perceived appreciation. Note that the memory R_m index did take these into account, which together with the use of a good chromatic adaptation transform and uniform colour space might account for the R_m's better performance.

From Figure 4 and Table 1, it is clear that Thornton's *CPI* index has a performance comparable to that of the gamut- and chroma-enhancement based metrics (Q_p, Q_g, *FCI* and *GAI*) and to the mixed $GAIR_a$ metric. This was confirmed by the CI-test (see Table 2) and largely by the MDS test which showed slightly lower performance, i.e. larger metric-rating distances, for the *FCI* and *GAI* metrics.

In agreement with the discussion on the interpretation of the first axis of the MDS analysis, it can also be observed that the predictive performance for visual appreciation tends to drop the less emphasis a metric places on chroma enhancement or gamut expansion. For example, the Judd flattery index, which uses only one-fifth of the preferred colour shift (primarily directed towards increased saturation), tends to have a lower performance compared to the Q_p, Q_g, *FCI*, *GAI* and $GAIR_a$ metrics. On the other hand, compared to the fidelity metrics, the R_f performance tends to be higher. As fidelity metrics do not and are not intended to emphasize chroma enhancement – except perhaps implicitly through the use of saturated reflectance samples – the former results are again consistent with the hypothesized role of chroma enhancement emphasis on predictive performance.

As a group, the fidelity metrics had the lowest performance of all (with the exception of Sanders' R_p). The performance of the CIE R_a and CQS Q_f were significantly lower than that of all the other metrics.

Finally, the role of chroma enhancement emphasis on predictive performance is also clearly illustrated by the four CQS indices, which show a statistically significant increase in performance as the reward for chroma enhancement in the index calculation is increased: $r_{Qf} < r_{Qa} < r_{Qp} \approx r_{Qg}$. Obviously, there will be an upper limit to the visually allowed chroma enhancement, as oversaturation is known to have a negative impact on perceived colour quality. Regarding the gamut-expansion (*FCI*, *GAI* and Q_g) and chroma enhancement (Q_a, Q_p and $GAIR_a$) based metrics, the former fail to account for such a limit. On the other hand, the latter and metrics such as R_p, R_f, *CPI*, R_m and $GAIR_a$ do include such a limit: either by setting up a reference chromaticity with increased saturation (e.g. R_p, R_f, *CPI* and R_m), or by explicitly setting a limit to the allowed chroma enhancement (e.g. Q_a and Q_p), or by the implicit counterbalance introduced by the changes in hue that are generally associated with increases in saturation (all former metrics, but especially $GAIR_a$).[68] In fact, the $GAIR_a$ counter-balances increase in gamut area (cfr. GAI) with full colour differences with respect to unsaturated reference chromaticities (cfr. CIE R_a). Therefore, increases in chroma or gamut are always associated with increases in colour difference.

An estimate of the upper limit for visually allowed chroma enhancement was obtained by optimizing the maximum allowed chroma enhancement in the CQS metric, $\Delta C*_{ab,max}$, for maximum Q_p performance. A maximum correlation ($r_{c,optimum} = 0.85$) was found for $\Delta C*_{ab,max} = 20$, which is twice the original value of 10 as set in *CQS v9.0*.

3.3.2. Naturalness

With the exception of Sanders' R_p index, all memory or preferred colour based metrics had approximately the same moderate correlation to naturalness ($r_c \approx 0.71$) as evaluated by the test subjects in the visual experiments. However, it should be noted that the MDS analysis did show some differences for these metrics. Although, the correlation of Sanders' R_p had increased for naturalness, it was still very poor ($r_c \pm 1SE = 0.16 \pm 0.21$). Obviously, sample mismatch and an outdated chromatic adaptation transform will have had a similar negative impact on predictive ability as before, so these will not be discussed further.

Compared to the results for visual appreciation, the performance of the memory R_m ($r_c \pm 1SE = 0.71 \pm 0.11$) and the *CPI* ($r_c \pm 1SE = 0.70 \pm 0.09$) had decreased, while Judd's flattery index had increased ($r_c \pm 1SE = 0.72 \pm 0.08$). The decrease of the former two can be explained by the idealized (and more chromatic) nature of memory and preferred colours. The increase of Judd's flattery index R_f can be explained by its use of less saturated reference chromaticities (R_f uses only one-fifth of the preferred colour shift).

That visual ratings on naturalness correlate better with the predictions of metrics that employ less saturated reference chromaticities is also indicated by the increase from low to moderate performance of the fidelity metrics CIE R_a ($r_c \pm 1SE = 0.60 \pm 0.12$), CRI2012 $R_{a,2012}$ ($r_c \pm 1SE = 0.61 \pm 0.08$) and CQS Q_f ($r_c \pm 1SE = 0.63 \pm 0.12$). The significantly higher correlation still of the CQS Q_a ($r_c \pm 1SE = 0.70 \pm 0.13$) and Judd's flattery index R_f ($r_c \pm 1SE = 0.72 \pm 0.11$) also show that the reference chromaticities supplied by illuminating a set of samples with a black-body radiator or daylight phase are not the most optimal in terms of naturalness (and especially not for visual appreciation). Naturalness seems to require a slight increase in saturation, in agreement with the interpretation of the first axis of the MDS analysis. As already argued in Davis and Ohno,[68] at the illumination level of indoor lighting chroma-enhanced rendering of object colours would make objects appear more like they would when illuminated by daylight (where the Hunt effect would lead to an increased object saturation). Obviously, there is a limit to the allowed saturation increase before performance starts to decrease, and as suggested by the discussion earlier and the drop in performance of the *CPI*, R_m and the strict gamut-/chroma-enhancement metrics (Q_p, Q_g, *FCI* and *GAI*) the limit for naturalness is smaller than that for visual appreciation. An estimate of the maximum visually allowed chroma enhancenment was obtained as before: $r_{c,optimum} = 0.82$ for $\Delta C*_{ab,max} = 7.5$ which is slightly smaller than the original value of 10 as set in *CQS v9.0* and about three times smaller than that obtained for visual appreciation.

According to both the correlation meta-analysis and the MDS analysis, the metric that performed best for naturalness was $GAIR_a$ ($r_c \pm 1SE = 0.94 \pm 0.09$). This metric is calculated as the geometric mean of the CIE R_a and the *GAI* metrics, thereby striking a balance between colour rendering fidelity and gamut expansion, without however placing boundary values on the individual metric scores as proposed by Freyssinier-Nova and Rea[70] in their statement that high values for both metrics ($R_a \geq 80$ and $80 \leq GAI \leq 100$) ensure "*positive subjective impressions of naturalness*". Although categorization makes decisions regarding the level of colour rendition easy, it is not what correlation analysis – which seeks to quantify the strength of the relationship between variables – is about. In addition, the requirement that the *GAI* should have values between 80 and 100 may be too strict, as there are quite a few light sources with good colour rendition that have *GAI* values larger than 100.[51]

4. Conclusions

Four colour rendition metrics based on memory or preferred colours are discussed (in chronological order): Sanders' preferred colour metric R_p, Judd's flattery index R_f, Thornton CPI and Smet's MCRI R_m. Sanders' and Smet's proposals are what can be termed pure memory/preferred colour rendition metrics, i.e. they use the actual memory or preferred colours in their index calculation. In addition, both account for the inherent psychophysical differences between chroma and hue tolerances by using a Mahalanobis distance, instead of the common Euclidean distance used in most colour difference equations. Judd's flattery index R_f and Thornton's *CPI,* on the other hand, compare the chromaticity of a number of Munsell samples illuminated by the test light source and the reference illuminant. The chromaticity under the reference illuminant is corrected by a preferred colour shift calculated by Judd based on memory and preferred colour data obtained in psychophysical experiments. While Thornton kept the original magnitude of the calculated preferred colour shift, Judd's flattery index rescaled it to one-fifth of its original length. Colour differences were calculated using a Euclidean distance metric, ignoring differences in chroma and hue tolerance in the assessment of the colour rendition of a light source.

The performance of these metrics was investigated, along with that of several other colour rendition metrics using psychophysical data on visual appreciation and naturalness obtained from, respectively, 21 and 15 experiments described in literature. Performance was analysed as the weighted average artefact-corrected Spearman correlation coefficient determined according to the method of Hunter and Schmidt[71] and an MDS based metric-rating distance.

Regarding visual appreciation, the R_m metric was found to correlate highly and significantly better than all other metrics. The good and best performance of the R_m metric was also confirmed by the results of the MDS analysis.

The other preferred colour-based metrics – Sanders' R_p, Judd's flattery index R_f and Thornton *CPI* – showed, respectively, very poor, poor-to-moderate and moderate-to-good predictive performance in terms of visual appreciation. Possible reasons for the lower performance of these three metrics were identified as being: The mismatch between metric sample set and the experiment objects set in the case of Sanders' $R_{p,}$ the use of only one-fifth of the preferred colour shift in Judd's flattery index, the use of a non-uniform colour space and outdated chromatic adaptation transform, and the use of a Euclidean colour difference equation that does not take differences in chroma and hue tolerance into account.

The chroma enhancement based metrics (Q_p, Q_g, *GAI*, *FCI* and the mixed $GAIR_a$) had moderate to good performance, comparable to that of the *CPI*.

The fidelity-type metrics had the worst performance (with the exception of Sanders' R_p), indicating the CIE defined reference illuminants are not optimal for predicting visual appreciation. The poor performance of the fidelity-type metrics is not surprising as they were never intended to predict these subjective aspects of colour rendition. Their only goal is to provide an objective measure to compare colour rendering (fidelity) of different light sources.

Considering the calculation details and performance of each metric, it was concluded that the more emphasis a metric places on chroma enhancement (or gamut expansion) the better its predictive performance in terms of visual appreciation tends to be. This is obviously only valid up to a certain limit, as oversaturation will have a negative impact.

For naturalness, with the exception of Sanders' R_p, all memory/preferred colour-based metrics performed approximately equally moderately according to the correlation analysis. However, the MDS analysis did suggest there were differences in performance: $R_p < R_m < CPI < R_f$.

The metric that performed best was the $GAIR_a$. The metric is calculated as the geometric mean of the *GAI* and the *CIE* R_a values. No significant difference could be found between the $GAIR_a$ metric and Judd's flattery index and the strict gamut-/chroma-enhancement metrics.

Compared to their performance for visual appreciation, the performance of the fidelity-type metrics for naturalness had increased from poor to moderate. That of the Q_a metric and R_f metric had increased even higher, while the R_m, *CPI* and the Q_p had all decreased to a rather moderate performance. The pure gamut area-based metrics, such as the *GAI*, Q_g and *FCI* had decreased to substantially lower performance levels.

Considering the observed trends in performance and the emphasis each metric places on gamut expansion or chroma enhancement, it was concluded that naturalness requires higher object saturation levels than the ones provided by the CIE reference illuminants in fidelity-type metrics, but not as high as for visual appreciation.

Funding

Author K Smet was supported through a Postdoctoral Fellowship of the Research Foundation Flanders (FWO) (12B4913N).

References

1 Hering KE. *Grundzüge der Lehre vom Lichtsinn*. Berlin: Springer-Verlag, 1920.
2 Bartleson CJ. Memory colors of familiar objects. *Journal of the Optical Society of America* 1960; 50: 73–77.
3 Bartleson CJ. Color in memory in relation to photographic reproduction. *Photographic Science and Engineering* 1961; 5: 327–331.
4 Bartleson CJ, Bray CP. On the preferred reproduction of flesh, blue-sky, and green-grass colors. *Photographic Science and Engineering* 1962; 6: 19–25.
5 Bodrogi P, Tarczali T. MacDonald LW, Luo MR. Investigation of colour memory. *Colour Image Science: Exploiting Digital Media*. Chichester: John Wiley and Sons Limited, 2002: pp. 23–48.
6 Bruner JS, Postman L, Rodrigues J. Expectation and the perception of color. *American Journal of Psychology* 1951; 64: 216–227.
7 de Fez MD, Capilla P, Luque MJ, Perez-Carpinell J, del Pozo JC. Asymmetric colour matching: Memory matching versus simultaneous matching. *Color Research and Application* 2001; 26: 458–468.

8 Duncker K. The influence of past experience upon perceptual properties. *American Journal of Psychology* 1939; 52: 255–265.
9 Fernandez S, Fairchild MD, Braun K. Analysis of observer and cultural variability while generating "preferred" color reproductions of pictorial images. *Journal of Imaging Science and Technology* 2005; 49: 96–104.
10 Granzier JJM, Gegenfurtner KR. Effects of memory colour on colour constancy for unknown coloured objects. *Perception* 2012; 3: 190–215.
11 Hansen T, Olkkonen M, Walter S, Gegenfurtner KR. Memory modulates color appearance. *Nature Neuroscience* 2006; 9: 1367–1368.
12 Hurlbert AC, Ling Y. If it's a banana, it must be yellow: The role of memory colors in color constancy. *Journal of Vision* 2005; 5: 787.
13 Kanematsu E, Brainard DH. No measured effect of a familiar contextual object on color constancy. *Colour Research and Application* 2014; 39: 347–359.
14 Ling Y. *The colour perception of natural objects: familiarity, constancy and memory*. PhD thesis. Newcastle upon Tyne: University of Newcastle, 2005.
15 Park D-S, Kwak Y, Ok H, Kim CY. Preferred skin color reproduction on the display. *Journal of Electronic Imaging* 2006; 15: 041203–041209.
16 Pérez-Carpinell J, de Fez MD, Baldoví R, Soriano JC. Familiar objects and memory color. *Color Research and Application* 1998; 23: 416–427.
17 Saito M. Comparative studies on color preference in Japan and other Asian regions, with special emphasis on the preference for white. *Color Research and Application* 1996; 21: 35–49.
18 Sanders CL. Colour preferences for natural objects. *Journal of the Illuminating Engineering Society* 1959; 54: 452–456.
19 Schloss KB, Strauss ED, Palmer SE. Object color preferences. *Color Research and Application* 2013; 38: 393–411.
20 Siple P, Springer RM. Memory and preference for the colors of objects. *Perception and Psychophysics* 1983; 34: 363–370.
21 Smet KAG, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. Colour appearance rating of familiar real objects. *Color Research and Application* 2011; 36: 192–200.
22 Tarczali T, Park D-S, Bodrogi P, Kim CY. Long-term memory colors of Korean and Hungarian observers. *Color Research and Application* 2006; 31: 176–183.
23 Yano T, Hashimoto K. Preference index for Japanese complexion color under illumination. *Journal of Light and Visual Environment* 1998; 22: 54.
24 Yendrikhovskij SN, Blommaert FJJ, de Ridder H. Representation of memory prototype for an object color. *Color Research and Application* 1999; 24: 393–410.
25 Zeng H, Luo R. *Modelling memory colour region for preference colour reproduction*: *SPIE: Color Imaging XV: Displaying, Processing, Hardcopy, and Applications*, 2010.
26 Yendrikhovskij SN, Blommaert FJJ, de Ridder H. Color reproduction and the naturalness constraint. *Colour Research and Application* 1999; 24: 52–67.
27 Bodrogi P, Tarczali T. Colour memory for various sky, skin, and plant colours: Effect of the image context. *Color Research and Application* 2001; 26: 278–289.
28 Xue S, Tan M, McNamara A, Dorsey J, Rushmeier H. *Exploring the use of memory colors for image enhancement*: *SPIE: Human Vision and Electronic Imaging XIX*, 2014.
29 Judd DB. A flattery index for artificial illuminants. *Journal of the Illuminating Engineering Society* 1967; 62: 593–598.
30 Sanders CL. Assessment of color rendition under an iIlluminant using color tolerances for natural objects. *Journal of the Illuminating Engineering Society* 1959; 54: 640–646.
31 Smet KAG, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. Memory colours and colour quality evaluation of conventional and solid-state lamps. *Optical Express* 2010; 18: 26229–26244.
32 Smet KAG, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. A memory colour quality metric for white light sources. *Energy and Buildings* 2012; 49: 216–225.

33 Thornton WA. A validation of the color-preference index. *Journal of the Illuminating Engineering Society* 1974; 4: 48–52.
34 Jost-Boissard S, Fontoynont M, Blanc-Gonnet J. Thurstone scalings for attractiveness of fruit and vegetables under nine 3000 K and eight 4000 K light sources. (personal communication) 2009.
35 Jost-Boissard S, Fontoynont M, Blanc-Gonnet J. Perceived lighting quality of LED sources for the presentation of fruit and vegetables. *Journal of Modern Optics* 2009; 56: 1420–1432.
36 Narendran N, Deng L. *Color rendering properties of LED light sources: Solid State Lighting II: Proceedings of SPIE*, 2002.
37 Rea MS, Freyssinier JP. Color rendering: Beyond pride and prejudice. *Color Research and Application* 2010; 35: 401–409.
38 Rea MS, Freyssinier-Nova JP. Color rendering: A tale of two metrics. *Color Research and Application* 2008; 33: 192–202.
39 Schanda J, Madár G. *Light source quality assessment: Proceedings of the CIE 26th Session*, Beijing, Vienna: CIE, 2007, pp.D1-72–75.
40 Szabó F, Csuti P, Schanda J. *Colour preference under different illuminants – new approach of light source colour quality: Light and Lighting Conference with Special Emphasis on LEDs and Solid State Lighting*, 27–29 May, Budapest, Hungary. Vienna: CIE. PWDAS-43.
41 Vanrie J. Appendix 4: Technical report to the user committee of the IWT-TETRA project (80163): The effect of the spectral composition of a light source on the visual appreciation of a composite objectset. Diepenbeek, Belgium: PHL, 2009 June 11, 2009. Report No.
42 Tsukitani A, Lin Y. *Research on FCI: Presented at the meeting of TC1-91*, Kuala Lumpur, Malaysia, April 23–26: 2014.
43 Imai Y, Kotani T, Fuchida T. *A study of color rendering properties based on color preference in adaptation to LED lighting: CIE2012 Lighting Quality & Energy Efficiency*. Hangzhou, China. Vienna: CIE, 2012, pp.369–374.
44 Houser KW, Tiller DK, Hu X. Tuning the fluorescent spectrum for the trichromatic visual response: A pilot study. *Leukos* 2005; 1: 7–23.
45 Smet KAG, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. Optimization of colour quality of LED lighting with reference to memory colours. *Lighting Research and Technology* 2012; 44: 7–15.
46 Ohno Y, Davis W. Visual evaluation experiment on chroma enhancement effects in color rendering of light sources, report submitted to TC1-69. National Institute of Standards and Technology, 2010.
47 Wei M, Houser KW, Allen GR, Beers WW. Color preference under LEDs with diminished yellow emission. *Leukos* 2014; 10: 119–131.
48 Imai Y, Kotani T, Fuchida T. *A study of color rendering properties based on color preference of objects in adaptation to LED lighting: CIE Centenary Conference "Towards a New Century of Light"*, Paris, France, Vienna: CIE, 2013, pp.62–67.
49 Tsukitani A, editor. *Optimization of colour quality for landscape lighting based on feeling of contrast index: CIE Centenary Conference "Towards a New Century of Light"*. Paris, France, Vienna: CIE. 2013: 68–71.
50 Smet K, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. Correlation between color quality metric predictions and visual appreciation of light sources. *Optics Express* 2011; 19: 8151–8166.
51 Houser KW, Wei M, David A, Krames MR, Shen XS. Review of measures for light-source color rendition and considerations for a two-measure system for characterizing color rendition. *Optics Express* 2013; 21: 10393–10411.
52 Guo X, Houser KW, Akashi Y. A review of colour rendering indices and their application to commercial light sources. *Lighting Research and Technology* 2004; 36: 183–199.
53 Boyce P, Smet K. LRT symposium 'Better metrics for better lighting' – a summary. *Lighting Research and Technology* 2014; 46: 619–636.
54 Houser KW, Mossman MA, Smet KAG, Whitehead L. Tutorial: Color rendering and its applications in lighting. *Leukos* 2015; accepted for publication.
55 Smet KAG, Lin Y, Nagy BV, Németh Z, Duque-Chica GL, Quintero JM, Chen H-S,

Luo RM, Safi M, Hanselaer P. Cross-cultural variation of memory colors of familiar objects. *Optics Express* 2014; 22(26): 32308–32328.

56 Commission Internationale de l'Eclairage. CIE13-1965. *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*. Paris, France: CIE, 1965.

57 Nickerson D, Jerome CW. Color rendering of light sources: CIE method of specification and its application. *Illuminating Engineering* 1965; 60: 262–271.

58 Newhall SM, Burnham RW, Clark JR. Comparison of successive with simultaneous color matching. *Journal of the Optical Society of America* 1957; 47: 43–54.

59 Commission Internationale de l'Eclairage. CIE13.3-1995. *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*. Vienna, Austria: CIE, 1995.

60 Smet KAG, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. Colour appearance rating of familiar real objects. *Color Research and Application* 2011; 36: 192–200.

61 Vurro M, Ling YZ, Hurlbert AC. Memory color of natural familiar objects: Effects of surface texture and 3-D shape. *Journal of Vision* 2013; 13: 20.

62 Ebner F, Fairchild MD. *Development and testing of a color space (IPT) with improved hue uniformity*: *Proceedings of the IS&T 6th Color Imaging Conference,* Scottsdale, Arizona, USA, 1998: 8–13.

63 Breneman EJ. Corresponding chromaticities for different states of adaptation to complex visual fields. *Journal of the Optical Society of America A – Optics Image Science and Vision* 1987; 4: 1115–1129.

64 Kuriki I, Uchikawa K. Adaptive shift of visual sensitivity balance under ambient illuminant change. *Journal of the Optical Society of America A – Optics Image Science and Vision* 1998; 15: 2263–1174.

65 Fairchild MD. Formulation and testing of an incomplete-chromatic-adaptation model. *Color Research and Application* 1991; 16: 243–250.

66 Smet K, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. A memory colour quality metric for white light sources. *Energy and Buildings* 2012; 49: 10.

67 Smet KAG, Schanda J, Whitehead L, Luo RM. CRI2012: A proposal for updating the CIE Colour Rendering Index. *Lighting Research and Technology* 2013; 45: 689–709.

68 Davis W, Ohno Y. Color quality scale. *Optical Engineering* 2010; 49: 033602–033616.

69 Hashimoto K, Yano T, Shimizu M, Nayatani Y. New method for specifying color-rendering properties of light sources based on feeling of contrast. *Color Research and Application* 2007; 32: 361–371.

70 Freyssinier-Nova JP, Rea MS. *A two-metric proposal to specify the color-rendering properties of light sources for retail lighting: Tenth International Conference of Solid-State Lighting, Proceedings of SPIE*, San Diego, USA, August 1–5: 2010, pp.77840V.

71 Hunter JE, Schmidt FL. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. 2nd Edition, Newbury Park, CA: Sage Publications, 2004.

72 Sanchez-Meca J, Marin-Martinez F. Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods* 2008; 13: 31–48.

73 Garcia PA, Huertas R, Melgosa M, Cui G. Measurement of the relationship between perceived and computed color differences. *Journal of the Optical Society of America A* 2007; 24: 1823–1829.

74 Zou GY. Toward using confidence intervals to compare correlations. *Psychological Methods* 2007; 12: 399–413.

75 Csuti P, Schanda J. Colour matching experiments with RGB-LEDs. *Color Research and Application* 2008; 33: 108–112.