Received: 4 September 2019 | Revised: 7 May 2020 | Accepted: 8 May 2020

DOI: 10.1002/col.22525

RESEARCH ARTICLE

WILEY

Impact of the adapted white point and the cultural background on memory color assessments

Sebastian Babilon | **Tran Quoc Khanh**

Technische Universität Darmstadt, Fachgebiet Lichttechnik, Darmstadt, Germany

Correspondence
Sebastian Babilon, Technische Universität Darmstadt, Fachgebiet Lichttechnik, Hochschulstr. 4a, Darmstadt 64289, Germany.
Email: babilon@lichttechnik.tu-darmstadt.de

Abstract

With their inherent ability of serving as an internal reference, memory colors provide a very powerful concept in the evaluation of color rendering properties of white light sources with respect to visual appreciation. Recent results for example suggest fairly good correlations between memory-based color quality metrics and the observers' general color preferences. However, due to technical limitations in the design of the underlying psychophysical experiments, they generally lack the explicit inclusion of realistic viewing and adaptation conditions, which is supposed to have a nonnegligible impact on the model prediction performance. In addition, intercultural effects might play a crucial role in the context of memory colors. For these reasons, the current article investigates the impact of both the adapted white point and the observers' cultural background on memory color assessments in order to contribute to a better understanding of these dependencies and their interactions. For this purpose, the color appearance rating results of Chinese and German observers were collected for a selection of 12 different familiar test objects assessed under two different adaptation conditions at 3200 K and 5600 K, respectively. From the statistical analysis of the experimental data, it is shown, in accordance to previous studies, that the impact of the observed intercultural deviations is likely to be of no practical importance even though significance is found. Despite considerably larger effect sizes, the same must be concluded for the two tested adaptation conditions.

KEYWORDS

color appearance, color rendering, color vision, memory colors, perception psychology, vision adaptation

1 | INTRODUCTION

With the advent of fluorescent gas-discharge lamps and, at a later stage, of light emitting diodes (LEDs), a completely new spectral flexibility in designing light emitting devices was introduced, which made developers and manufacturers strive for new tools and algorithms for tweaking the light emission of an illuminant toward high user acceptability and preference. In general, preference in lighting applications is always a matter of how appealing the objects in an illuminated scenery appear to the individual observer, which is a very subjective aspect

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. *Color Research and Application* published by Wiley Periodicals LLC

of color quality and, therefore, makes a comprehensive description of the associated phenomena a very challenging task. On the other hand, it would be favorable to have a universally valid, single-color quality metric, which can be used in a large variety of different situations to appraise whether the light emitted by an illuminant is judged to be of high or poor quality when assessed by an observer. In the past, the Commission Internationale de l'Éclairage (CIE) as well as a large number of research groups all over the globe addressed this topic—often with rather moderate success.[1-3]

However, in this context, the most promising approaches in specifying the perceived color quality of light sources in real-lighting applications are those based on internal references such as preferred or memory colors. The reason for this is that in the absence of an external reference, that is, a reference illuminant, daylight, and so on, which is usually the case for the majority of lighting applications, object colors and, consequently, the color quality of an illumination are often judged in relation to what people expect the perceived objects to look like.[4-9] Depending on the specific application, this could be either a question of preference, of naturalness, or of any other subjective aspect of color quality. In any case, it always involves the comparison of the perceived scene with an inherent reference, which generally has to be recalled from memory.

In order to be able to appropriately model such memory-based color assessment over a wide range of different lighting situations, it is important to know the potential impact factors and to understand the way they influence people's inherent references, that is, their memory colors. In this article, two different potential impact factors will be investigated: (a) Intercultural differences between individual observer groups and (b) variations in the chromaticity of the adapted white point expressed in terms of correlated color temperatures (CCTs). The intent of the underlying research effort is basically 2-fold in such that the key results and findings of previous studies on the assessment of memory colors, in particular with regard to crosscultural variations, should be verified and, at the same time, extended to somewhat more application-related immersive viewing and adaptation conditions.

For this purpose, the article is organized as follows. Section 2 starts with a short discussion of previous studies dealing with the assessment of memory colors and their limitations. In Section 3, a short overview of the experimental setup of the current work, the corresponding test method and the way of modeling the observers' color appearance ratings for different cultural observer groups and illumination conditions will be presented. In Section 4, the results and general characteristics of these memory color rating experiments will be discussed. Based on this descriptive evaluation, Sections 5 and 6 are dedicated to the statistical analysis of the color appearance rating data in order to determine whether or not the observers' state of chromatic adaptation and their cultural background have a significant impact on the memory color representations. Finally, a short summary of the key findings of this work as well as an outlook on future research intentions will be given in Section 7.

2 | PREVIOUS STUDIES

In the past, several different studies were published trying to specify color rendering properties of white light sources by adopting a single preference- or memory-based color quality metric.[1,6,10-14] Even though some of them offer an excellent methodology, technical limitations can be identified in the reported experimental designs.

One of these limitations with regard to real-world lighting scenarios is that previous work generally does not take into account the influence of realistic adaptation and viewing conditions on the observer ratings. In particular, it was shown that object naturalness, such as realistic texture, shading, motion parallax, and binocular disparity, has a provable impact on similarity ratings.[15,16] Thus, an appropriate selection of real-test objects and their assessment under immersive viewing conditions might be beneficial for a metric's predictive performance in real-lighting applications when being derived from the outcome of such more realistic experiments.

Furthermore, color perception and, therefore, the perceived color appearance of the test objects vary substantially with the observers being adapted to different viewing environments[17-19] so that preference in lighting color quality studies usually depends on the specific context.[20] Contextless viewing cabinets or viewing-booth-like approaches, which were often applied in previous studies, possibly cannot suitably account for these effects. On the other hand, though, one may argue that the use of such contextless cabinets might actually be beneficial when deriving memory colors for a general color rendition index, as they do not bias the observers with a specific context. With both arguments appearing reasonable from their individual point of view, it is therefore important to find the right balance in the experimental design of providing immersive and realistic viewing conditions while still keeping the setup as general as possible.

Regarding the two most prominent studies on the assessment of memory colors—one was conducted by Sanders[5] in the late 1950s, the other by Smet et al[9] published in 2011—the following limitations can be identified. First of all, as discussed above, they both lack realistic viewing and adaptation conditions, that is, the test objects were presented to the observers through a small quadratic aperture of a viewing cabinet. Even though Smet et al in contrast to Sanders attempted to ensure more general viewing conditions and a fixed adaptation state by extending the adaptation conditions of the inside of the box to the outside, their experimental setup is still far from representing a realistic viewing environment. Second, both former studies' object selection in general lacks of hue coverage regarding the more saturated parts of the hue circle. As stated by Davis et al,[21,22] this could be problematic for the predictive performance of a color quality metric based on such low to medium saturated color samples. In particular, the peaked spectra of white LED light sources could be optimized for a high metric value even though the actual perceived color quality is much poorer. By the inclusion of higher saturated color samples in the metric definition, the risk of such malpredictions would be reduced to a minimum.[3] Third and last, it must be discussed, with regard to future perspectives, whether a CIECAM02-based color space might be a better choice for the current experiments to model observers' memory color appearance ratings than the color spaces used by Sanders and Smet et al, respectively.

While Sanders, for historic reasons, adopted the anything but perceptually uniform CIE 1931 (x,y) chromaticity space[23] in conjunction with an outdated and poorly performing translational chromatic adaptation transform,[24] Smet et al employed the more sophisticated IPT color space,[25,26] which was additionally extended by the latest CAT02 transform. This allowed, in contrast to the method used by Sanders, for an approximately correct prediction of the impact of the observers' state of chromatic adaptation. Further reasons for choosing the IPT color space were its capability of modeling color appearance in the presence of a self-luminous background, which was the case in their experimental design, as well as its sufficiently good hue uniformity.[27] The latter was considered as an important feature since most of the rating distributions reported by Smet et al[9] were oriented along the radial direction indicating that observers were less tolerant of shifts in hue than of shifts in chroma.

Even though CIECAM02-based color spaces are found to exhibit somewhat larger hue errors than the IPT color space when evaluating hue uniformity, they perform significantly better when it comes to the evaluation of color differences. As shown by Luo et al,[28] this holds true for large but even more for small color differences. In addition, CIECAM02 has the CAT02 natively included and also considers color appearance phenomena such as the Hunt, Stevens, surround, and lightness contrast effects. The use of a CIECAM02-based color space might therefore be beneficial with regard to future applications.

Regarding the effect of crosscultural variations, two different studies—one conducted by Tarczali et al,[29] the other by Smet et al[30]—were found in the literature focusing on the investigation of the characteristics of memory colors under that specific aspect. While Tarczali et al considered only two different regions by comparing memory color assessments of Hungarian and Korean observers, Smet et al performed a more extensive investigation including observers from seven different countries around the world. In both studies, care was taken that the same controlled display-based experimental conditions were used at all locations so that potentially occurring differences in the memory color assessments could directly be attributed to the cultural differences between the various observer groups rather than to variations in the experimental setup. Even though in both cases statistically significant differences were reported for the cultural results, Smet et al stated that the crossregional variations were found to be of the same order as or smaller than the interobserver variability within a single geographical region. Similar results were also found by Fernandez et al,[31] who performed the analysis of observer and cultural variabilities for the preferred color reproduction of pictorial images including four different cultural backgrounds.

From these findings, it was therefore concluded that the impact of the observers' cultural background is likely to be of little or no practical importance for the assessment of memory colors.[30] Due to the use of controlled experimental conditions for all participating regions, the transferability of the obtained results to more realistic viewing environments seems to be justified and should experimentally be validated by the current work. First indication to this transferability in a more implicit manner was also given by Smet and Hanselaer[32] who, based on the results from the previous study, defined regional-specific memory-based color quality measures that they eventually used to investigate the impact of crossregional differences on the evaluation of color rendition in terms of predictive performance. They found that even though small differences in the absolute level of color rendition were observed, the regional-specific measures were generally comparable in predicting light source rank order and correlating with visual data. In this context, Smet et al's MCRI[14] was found to be a good approximation to a universally valid global metric.

In a series of recently published studies,[33,34] memory colors were also used as internal references to investigate chromatic adaptation. In the corresponding experiments, memory color matching was applied on five different real-familiar test objects to derive sets of corresponding colors for several different illumination conditions. These sets, including the test objects' respective memory color representations, were then used to evaluate the performance of various linear and nonlinear chromatic adaptation transforms. The test objects were presented to the observers in a viewing-booth-like environment equipped with a background scene which was populated by several three-dimensional, gray-colored objects to enhance scene realism by providing depth, parity, and illumination cues. Even though a relatively large 50° field of view of the adapting (background) field was achieved, their experimental design still lacked immersive realistic viewing and, therefore, shows the same limitations as discussed above.

When being exposed to an immersive illumination condition in a real-sized room, observers must be assumed to be in a significantly different cognitive state compared to what is expected when they assess object representations and images viewed through a booth setup or while being seated in front of a computer monitor. Thus, with regard to the assessment of memory colors in real-world lighting scenarios, additional experiments adopting realistic viewing and adaptation conditions, such as the one presented in the following, are therefore required to verify the results and conclusions found in the literature with regard to a more application-related approach.

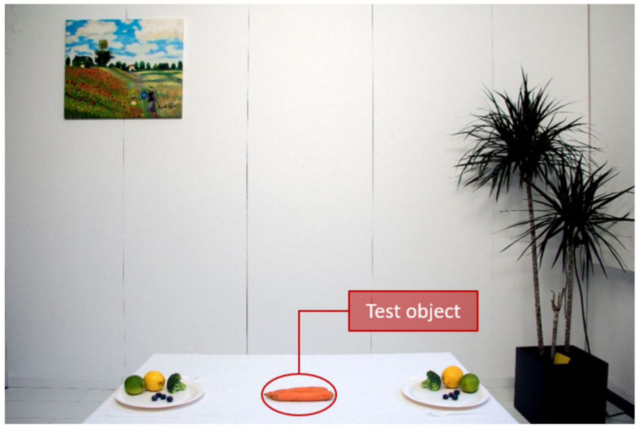

FIGURE 1 Experimental setup for investigating the cultural influence and the impact of the adapted white point on the color appearance rating of familiar real objects. The image shows an approximate representation of the observers' perspective when assessing the test objects. The light sources which are mounted to a truss hanging above the observers' heads and the couch the observers were seated on are not shown here

3 | EXPERIMENTAL METHOD AND TEST PROCEDURE

Based on the discussions of the previous section, a new experimental approach was developed by the authors and has been presented recently.[35] As stated in the introduction of this article, it can be considered as an extension of previous studies, in particular of those conducted by Sanders[5] and Smet et al,[9] to immersive and realistic viewing and adaptation conditions. Thus, many of their concepts were reused and adopted here. In contrast to their viewing-booth-like environments, however, the current approach requires the corresponding rating experiments to be conducted in a real-sized room. For the right balance in the experimental design, a clean, white surrounding was chosen with some additional small colored objects being placed in the peripheral field of view (see Figure 1). These objects are supposed to provide clues to the visual system for achieving color constancy[36,37] and keeping the observers' state of chromatic adaptation stable even when the test object's color appearance in the central field of view is changing throughout the experiments.

Originally, the experiments had been conducted with only a group of German observers adapted to a white point of 5600 K. For the current article, though, these experiments were repeated for a second cultural observer group of native Chinese participants and a second adaptation condition of 3200 K so that both the impact of the adapted white point and the observers' cultural background on the memory color assessments could be investigated. In the following, a brief description of the experimental setup will be given. For further details including additional image representations, the interested reader is referred to the original paper.[35]

3.1 | Test object selection and experimental setup

The object selection was based on an online-survey in which participants were asked to indicate object names that first came into their minds when thinking of a specific given color or, more precisely, hue region. In total, 1009 people took part in this survey, which was distributed in German language via the homepage of the Deutsche Lichttechnische Gesellschaft e.V. and by email. Further details and the corresponding results can be accessed and downloaded as supplementary material from our institutional webpage.[38] A brief description of the main results and evaluation procedure follows.

Generally speaking, objects that were most frequently associated with certain hue regions, based on the outcome of the survey, were eventually chosen as the prototypical memory color objects for that specific parts of the hue circle adopted for the experiments. Exceptions were made for the hue regions of orange, yellowish-green, and purple. For these hues, a preevaluation performed on the corresponding most frequently named test objects of orange, green apple, and aubergine revealed that their surface characteristics were not Lambertian enough, thus causing too much specular reflection. This might have disturbed the observers color appearance ratings during the main experiments. For this reason, it was decided to choose less frequently named objects of the same hue region, which did not show such perturbations on their surfaces, as the final test objects. This procedure resulted in the following object selection: banana, green salad, broccoli, blue jeans, blueberry, red cabbage, red rose, carrot, butternut squash, and a concrete flowerplot. In addition, two different kinds of human complexion, that is, Asian and Caucasian skin, were added, eventually yielding a set of 12 different familiar real-test objects well distributed around the hue circle.

For an estimation of their typical reflectance characteristics, representatives of all test objects were either purchased or, in the case of the skin colors, real-human models were recruited from our faculty staff (the same as were later used for the experiments). Their corresponding reflectance spectra were subsequently determined from spectral radiance measurements of the objects' surfaces compared to a white reflectance standard, both homogeneously illuminated by a temporally stable halogen light source. For each test object, several measurements of characteristic surface points were made and the results were averaged. The resulting spectra can be downloaded from our institutional webpage.[38]

For all test objects, test conditions, and the two observer groups, interobserver and intraobserver variabilities are found to be sufficiently small and in the range of the variabilities reported by Smet et al[9] (see Section 4.2). This means that (a) naturally occurring variations in the objects' appearance have only a relatively small impact on the observers' ratings so that different observers of the same cultural background generally have quite stable and consistent versions of the same object in mind when performing the rating experiments and (b) a comparable and sufficient degree of familiarity with the final test object selection can be assumed for both cultural observer groups.

Once the test object selection was completed, the lighting situations for running the experiments had to be defined. For the current article, two different CCTs of 3200 K and 5600 K were adopted. While the latter CCT was chosen for being able to compare the current results directly to those reported by Smet et al[9,14] (see Babilon and Khanh[35] for this comparison), the former represents the typical bias point of tungsten halogen fixtures used for theater and film productions as well as for lighting applications in the shopping and retail industry, that is, in fields of application where the aspirations of providing excellent light sources for high-quality lighting are traditionally strongly pronounced.[39]

In order to provide proper adaptation conditions at both CCTs, two dimmable four-channel LED panels of Lambertian light emission were used. Hanging from a truss placed at the back wall of the experimental room, their light emitting surfaces were oriented in such a way that a homogeneous illumination of whole room was guaranteed. For both CCTs, the horizontal illuminance at the test objects's position was approximately 2000 lx. Measurements of the luminances in the central and peripheral field of view revealed consistent values of about 500 cd/m^2 with an average uniformity of $E_{\mathrm{min}}/\bar{E}>0.82$. A similar luminance level of 590 cd/m^2 was reported by Smet et al[9] for their applied background adaptation condition, which was provided by a self-luminous white panel mounted to the back wall of their illumination box. Depending on the level of emitted radiance of the four LED channels used for shifting the test objects color appearance, this resulted in a corresponding vertical illuminance at the test object's position ranging from 430 lx (all LEDs turned off) to 1150 lx (all LEDs at maximum power).

The measured spectral power distributions of the light spectra used for the current experiments are depicted in Figure 2A,B, respectively. They were both optimized for excellent color rendition showing a general color fidelity index R_f of 92.8 for the 3200 K and 91.7 for the 5600 K case. The high illuminance level of approximately 2000 lx of the ambient light in combination with the relatively homogeneously illuminated viewing field (table plus opposing wall) and the small colored objects additionally placed on the table for increasing color constancy guaranteed a fast, stable, and complete chromatic adaptation of the observers during the experiments. This was empirically confirmed by the fact that for both adaptation conditions no remaining tint in the scene white could be perceived or reported by the observers after five minutes of initial adaptation.

Furthermore, the excellent agreement of the photometric quantities between both adaptation conditions precluded the adulterating influence of varying luminance levels on the observers' color appearance ratings in the current experimental setup. Hence, potentially occurring differences in the outcome of the experiments conducted at different adaptation conditions may solely be explained by the impact of the different adapted white points rather than by a change in any of the photometric quantities.

Finally, it should be noted that the reason for choosing considerably larger illuminance levels than usually found as general or recommended settings in indoor lighting is mainly 2-fold. First of all all, it has clearly been shown in the literature that when it comes to visual preference, which, by definition, goes beyond the simple

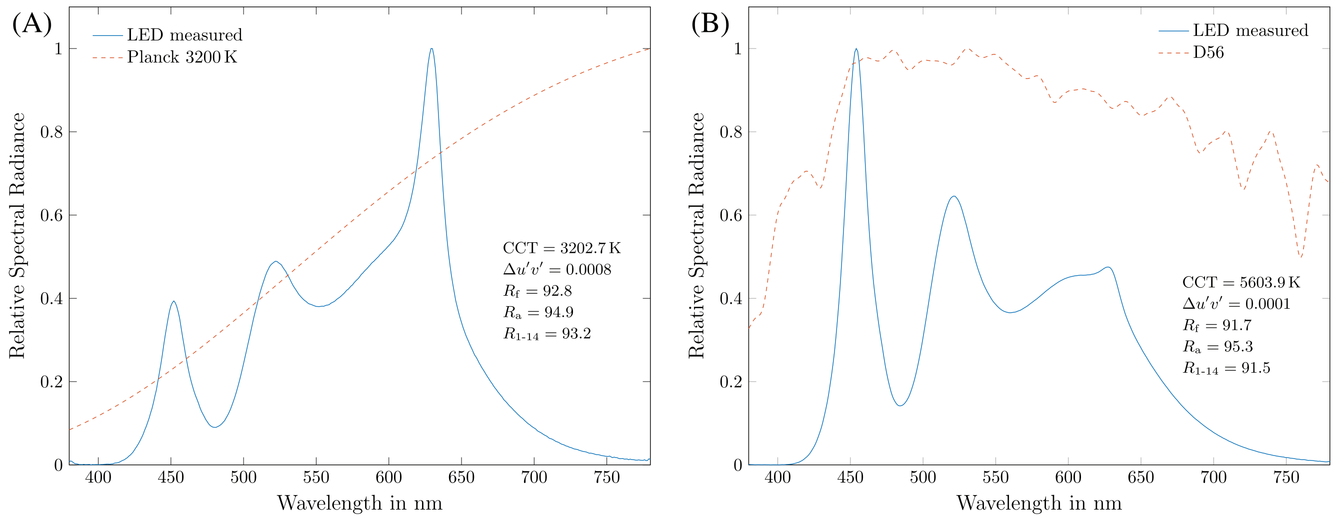

FIGURE 2 Normalized measured spectral radiance of the optimized four-channel LED light source for 3200 K (left) and 5600 K (right) ambient illumination determining the two different conditions of chromatic adaptation. Compared to the corresponding reference illuminants, that is, the Planckian radiator for the former and CIE D56 for the latter, excellent color rendition and white point preservation could be achieved

fulfillment of the visual task, the required illuminance must be significantly larger than what is usually reported in international standards on indoor lighting. In contrast to such standards, it is found that typical illuminance levels for interiors to achieve user preference and satisfaction should be of the order of 1300 lx to 2500 lx[40-46] (see also Khanh and Bodrogi[47] for a corresponding review).

Second, as shown by Witzel et al,[48] memory color effects are most strongly pronounced for objects whose typical colors were close to the daylight axis, which may be explained by the fact that the actual color of a particular object perceived in a more natural environment typically varies along this blueish-yellowish axis when perceived under different natural daylight illuminations. Based on this finding, which emphasizes the importance of natural daylight in the context of memory colors, it is likely that not just the adapted white point but also the illuminance level plays a crucial role regarding the assessment of memory colors of natural objects. With natural daylight usually providing high levels of illuminance, we therefore believe that performing memory color rating experiments at too low values would not adequately account for the memory color effects and the potential differences between different adaptation conditions one is actually interested in. Combining these two arguments, it can be concluded that performing our experiments at approximately 2000 lx is expected to be, as a starting point, a good trade-off between the requirements of user-preferred indoor lighting and the more fundamental deliberations on the nature of memory colors.

Besides providing the adaptation conditions under which the experiments should be conducted, an additional setup was required that allowed for shifting the chromatic appearance of the test objects while keeping their lightness components and, therefore, the observers' state of chromatic adaptation constant. For this purpose, an LCD projector was mounted next to the LED fixture, which, by masking each test object individually, was used for creating the illusion that the observed changes in color appearance were intrinsically caused by the objects themselves rather than by a change in the illumination keeping the observers' state of chromatic adaptation constant. A similar setup was applied successfully in several studies by Smet et al to investigate unique white settings[49,50] and chromatic adaptation using memory colors[33,34,51] as well as by Kraft and Brainard[52] to study the mechanisms of color constancy and by Ling and Hurlbert[16] to investigate the interactions between color and size in three-dimensional object similarity tasks.

For the rating process itself, approximately 65 color variations were realized per test object, which, from a colorimetric point of view, could be described as a two-dimensional grid of chromaticities in CAM02-UCS (cf Figure 3). This color representation space was selected here because (a) it is based on the CIECAM02 color appearance model with all its advantages in predicting the perceptional color appearance of object colors over a wide range of different viewing and adaptation conditions and (b) as shown by Luo et al,[28] outperforms most of the available color space or color difference formula alternatives, including CIELAB and IPT, in terms of predicting perceptual color differences, in which we are particularly interested in.

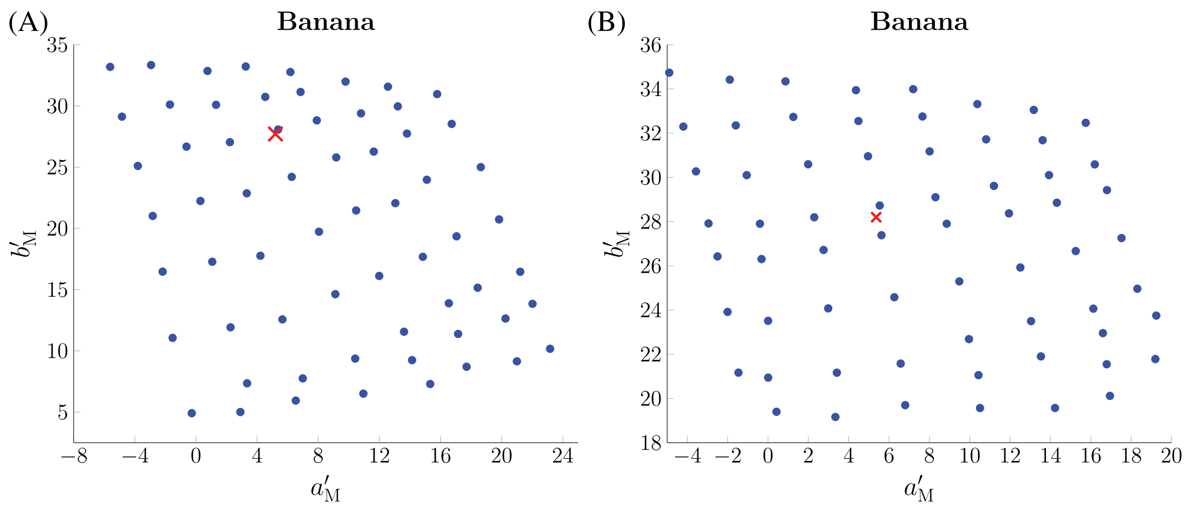

FIGURE 3 Exemplary CAM02-UCS chromaticity grids for the test object of banana as prepared for the test conditions at 3200 K (left) and 6500 K (right) ambient illumination. In each case, blue dots represent the object chromaticities that should be assessed by the observers, while the red cross indicates how the object would be perceived under the respective ambient illumination only

For each given stimulus defined by the corresponding CAM02-UCS grid point, the observers were asked to rate the similarity between the perceived color appearance of the respective test object and their idea of how the respective object looked like in reality on a semisemantic five-level scale of the Likert-type[53] running from "1" (very bad) to "5" (very good). Three subjects were tested at the same time, while they were advised to sit quite close to each other to get a similar view for properly rating the test object's color appearance. They were further instructed to keep their answers private and not to talk to each other during the time of the experiment. The average viewing distance to the test object lying on the coffee table in front of the subjects was about 1.5 m, resulting in a viewing angle of 5° to 20° depending on the size of the test object. This basically implies the use of the 10° color matching functions for all colorimetric calculations presented in this work including the calculation of the CAM02-UCS coordinates.

The order in which the different stimuli of a given test object were presented to the observers was completely randomized in all experimental sessions to average out potential learning and bias effects in the overall rating results. Each object stimulus was shown to the observers for exactly 12 seconds before switching to the next one so that it took the observers approximately 20 minutes to finish a single test run, which was supposed to be short enough to prevent fatigue.

3.2 | Modeling color appearance ratings

As shown by Yendrikhovskij et al,[15] the most appropriate way of modeling memory color assessments based on color appearance ratings sampled from a uniformly spaced grid (see Figure 3), is to fit for each test object a modified bivariate Gaussian distribution to the collected observer data. Besides giving a prototypical representation of the test object's memory color, this procedure also allows for quantifying the perceived similarity between the respective object's color appearance and the idea of how the object should look like in reality for an average observer. The latter, in particular, is considered to be an important feature for the derivation of a memory-based color-quality metric.[9]

Following the argumentation of Yendrikhovskij et al, similarity ratings, which are made for presented object colors selected from a uniformly spaced grid, sample the underlying similarity distribution. This distribution basically describes the likelihood that the perceived color belongs to a certain object category and, therefore, is proportional to the function value of the corresponding category distribution. Based on the assumption from general recognition theory that the structure of natural categories can effectively be modeled by using multivariate Gaussians, the similarity function can therefore be assumed to be consistent with a Gaussian distribution. Its centroid, in the context of memory colors, then gives the most likely color representation associated with the respective familiar test object.

Hence, bivariate Gaussian fitting was applied in the following to model for each test object the corresponding observer rating data of each cultural observer group as obtained for the two different adaptation conditions. In each case, the Gaussian fit functions were slightly modified in order to account for observer variability and observers not using the entire rating scale (cf Smet et al[9] and Babilon and Khanh[35]). This eventually results in four different sets of Gaussian fit functions describing for each of the two different cultural observer groups the prototypical memory color representations of the 12 familiar test objects assessed under both adaptation conditions at 3200 K and 5600 K, respectively. These sets of fit functions including both the goodness-of-fit measures as well as all model parameters such as the color center coordinates and the covariance matrices are available for download via our institutional webpage.[38]

3.3 | Observers

In total, 44 male and 37 female observers participated in the experiments. Most of them were recruited among the university students and faculty staff showing a varying degree of experience in color science. Table 1 summarizes their age and cultural structure. All observers were either native Chinese or native German people who had not been born or grown up abroad and had not been living outside of their home country for a longer period of time (>3 months). In order to avoid unwanted cultural bias for the Chinese observer group potentially induced by an advancing familiarization with German culture in the course of their time spent in Germany, care was taken to invite only those Chinese to participate in the experiments who had been living in Germany for less than 3 months. This was expected to be short enough so that their color appearance ratings could still be considered as "pure" Chinese.

Before being invited to take part in the experiments, all potential subjects were tested for normal color vision, which was done by using the Ishihara Test for Colour Deficiency,[54] the Standard Pseudoisochromatic Plates Part II for Acquired Color Vision Defects by Ichikawa et al,[55] and the Farnsworth-Munsell D-15 Color Vision Test.[56] Participants wearing glasses or contact lenses for visual acuity correction were not excluded from this study as long as the limits of ±6 dpt or 4 dpt in case of astigmatism were not exceeded. It was additionally ensured that all participants were familiar with the selected test objects. For the Chinese observers, it was also verified that they had already been familiar with these objects prior to their stay in Germany.

No further restrictions on the (random) appointment of individual participants regarding age range, gender, or experience in color science have been introduced, that is, all people that had been indicating interest in taking part in the experiments were eventually invited as long as they fulfilled the specific requirements on their cultural background and color vision abilities as discussed above. Even though this procedure led to an obvious deviation in the age ranges reported for the Chinese and German subjects, there is no significant difference in the group means of the age distributions. This is mainly due to the fact that only two German participants were older than 35, while the rest of the German observers showed a similar age structure as their Chinese counterparts. Hence, with no significant differences being revealed in the age distributions of the Chinese and German observer groups, it seemed to be justified to pool the corresponding within-group rating data without any further restrictions. The same holds true for the distributions of the individual experience levels which, being quantified based on a self-assessment questionnaire, were found to be equal in both cases. In addition, no discrepancies were observed in the color appearance ratings between younger (≤35 years) and older (>35 years) as well as between experienced and rather unexperienced observers resulting in quite consistent rating results among all subjects within each of the two different cultural observer groups.

Due to organizational and scheduling issues, each test object presented under the two adaptation conditions was rated by a total number of 15 observers per cultural group with an approximately equally balanced male-female ratio randomly picked (as good as possible) from the panel of subjects summarized in Table 1. Care was taken that all participants rated approximately the same amount of objects in order to reduce personal bias.

The reason here to focus on collecting the color appearance ratings of Chinese in comparison to German observers was their considerably different cultural background. The Chinese culture and traditions as well as the way how young people grow up in China can clearly be distinguished from what is usually experienced in Germany (or Central Europe in general). Chinese people for example eat different food, they live in a different natural and architectural environment, they get a different education, and so on. Due to all these reasons, it is assumed

TABLE 1 Overview of the age and cultural structure of the test panel of observers participating in the experiments

Cultural group	# Male	# Female	Range of ages	Average age
Chinese	18	19	19 to 35	25.2±2.9
German	26	18	19 to 64	26.5±8.9

Note: For each cultural group, the mean ages and corresponding ±1σ-intervals are tabulated in the last column.

that Chinese people could have developed their own understanding of what is beauty and what is pleasant to them that might be significantly different from what a German observer would expect. Hence, comparing these two observer groups seemed to be a good starting point for the investigation of the potential impact of the observers' cultural background on the memory color assessments in the current experimental design.

4 | COLOR APPEARANCE RATING RESULTS OF CHINESE AND GERMAN OBSERVERS

In the following, a descriptive analysis of the experimental data should be provided. For this purpose, chromatic differences of the memory color centers between Chinese and German observers as well as the corresponding tolerances in the observers' ratings are reported for each of the 12 different familiar test objects considered in this study. A comparison is made in terms of the two different adaptation conditions at 3200 K and 5600 K, respectively. In addition, interobserver and intraobserver performance factors are calculated to evaluate the observer variability in the memory color assessments of the present experiments.

4.1 | Chromatic differences and tolerances

Based on the similarity ratings of each cultural observer group, as described in Section 3.2, Gaussian modeling could eventually be applied to derive the memory color representations of an average Chinese or German observer for each of the 12 familiar test objects assessed under both adaptation conditions. For this purpose, following Smet et al,[30] the mean scores of the similarity ratings of both cultural observer groups obtained for each test object's respective set of CAM02-UCS $\left(a'_{\mathrm{M}}, b'_{\mathrm{M}}\right)$ chromaticity coordinates were fitted by bivariate Gaussian similarity distribution functions. From these model fits, the corresponding memory color centers were extracted as being given by the centroids of the Gaussian distributions and are summarized in Table 2. Additionally tabulated are the resulting differences in the memory color centers due to the different adaptation conditions (upper part of the table) and the observers' cultural background (lower part of the table).

Considering the former, deviations between the memory color centers obtained for the two different adaptation conditions ranging from 1.82 to 4.93 $\Delta E'_{\mathrm{chrom.}}$ with an average of 2.88 $\Delta E'_{\mathrm{chrom.}}$ and from 0.40 to 6.19 $\Delta E'_{\mathrm{chrom.}}$ with a slightly smaller average of 2.31 $\Delta E'_{\mathrm{chrom.}}$ can be found for the Chinese and German observer group, respectively. The indicated average deviations give a rough estimate for the order of magnitude of the effect of the adapted white point on the color appearance ratings. Even though this effect seems to be of the order of ~2.3 to 2.9 $\Delta E'_{\mathrm{chrom.}}$ for both cultural observer groups, it can however not yet be confirmed that the different adaptation conditions do have a significant impact. In addition to the descriptive analysis provided here, further statistical testing is necessary and will be performed as part of the discussions provided in Section 5.

By comparing, on the other hand, the Chinese memory color centers at both adaptation conditions with those obtained for the German observers as summarized in the lower part of Table 2, relatively moderate deviations between these two cultural observer groups can be found. For the 3200 K ambient illumination, chromatic differences ranging between 0.14 and 5.27 $\Delta E'_{\mathrm{chrom.}}$ with an average of only 1.85 $\Delta E'_{\mathrm{chrom.}}$ are observed, whereas the 5600 K adapted white point yields chromatic differences ranging from 0.23 to 2.69 $\Delta E'_{\mathrm{chrom.}}$ with an even smaller average of 1.19 $\Delta E'_{\mathrm{chrom.}}$. It should be noted that in both cases the observed chromatic differences between the two cultural observer groups are smaller than the average chromatic shift of the memory color centers induced by a change in the adaptation conditions. The corresponding statistical analysis will be provided in Section 6.

In order to visualize the chromatic differences and tolerances in the average observer's ratings of both cultural observer groups better, Figures 4 and 5 individually compare the acceptance boundary ellipses for each of the 12 familiar test objects that were assessed by the two different observer groups at 3200 K and 5600 K ambient illumination, respectively. These acceptance boundaries were essentially calculated from the fitted Gaussian distributions in such a way that their contour lines represent an average observer rating of "3," which is considered to be the just acceptable limit and, consequently, determines for each test object the chromatic tolerances observed in the current experiments. For convenience, the corresponding memory color centers are depicted as black crosses, while the dashed light-gray line segments illustrate the orientation of the ellipses' major semiaxes.

As can be seen, the characteristics of the fitted acceptance boundaries are strongly object-dependent and to some extent governed by the white point of adaptation. In most cases, the corresponding contour line plots exhibit similar shape and orientation but distinct differences in size when comparing the results of both cultural observer groups. With the exceptions of Asian skin, carrot, and green salad, the German observer ratings tend to give smaller tolerance ellipses for both adaptation conditions. This indicates that on average Chinese observers are slightly more tolerant of and, at the same time, less

TABLE 2 CAM02-UCS chromaticity coordinates a'_M and b'_M of the Chinese and German memory color centers which are obtained for the 12 familiar test objects assessed under both adaptation conditions at 3200 K and 5600 K ambient illumination

	Chinese observer group					German observer group				
	3200 K		5600 K			3200 K		5600 K		
Test object	a'_M	b'_M	a'_M	b'_M	$\Delta E'_{chrom.}$	a'_M	b'_M	a'_M	b'_M	$\Delta E'_{chrom.}$
Asian skin	12.92	10.16	10.29	10.86	2.72	13.04	10.09	10.29	10.36	2.76
Banana	3.88	28.29	4.44	30.83	2.60	5.07	31.77	5.08	32.59	0.82
Blueberry	−2.85	−12.81	−2.42	−10.54	2.31	−3.54	−12.62	−3.32	−12.29	0.40
Blue jeans	−7.63	−17.17	−7.12	−14.75	2.47	−8.11	−19.71	−7.79	−17.36	2.37
Broccoli	−9.62	14.38	−8.30	16.10	2.17	−8.62	9.20	−8.27	15.38	6.19
Butternut	15.75	20.64	14.19	19.70	1.82	14.73	20.47	15.38	20.57	0.66
Carrot	27.49	26.63	24.99	24.65	3.19	25.09	25.01	25.64	25.63	0.83
Caucasian skin	12.72	9.40	8.46	8.25	4.41	12.19	9.23	9.88	8.52	2.42
Concrete	−0.82	−3.54	−0.48	−1.35	2.22	−0.80	−3.33	0.08	−1.17	2.33
Green salad	−13.52	23.13	−9.56	26.06	4.93	−13.59	24.19	−10.43	26.70	4.04
Red cabbage	11.91	−16.93	10.41	−14.65	2.73	12.86	−13.65	10.56	−14.08	2.34
Red rose	37.11	13.21	34.89	11.25	2.96	36.66	12.75	34.68	11.12	2.56
	3200 K adapted white point					**5600 K adapted white point**				
	Chinese		**German**			**Chinese**		**German**		
Test object	a'_M	b'_M	a'_M	b'_M	$\Delta E'_{chrom.}$	a'_M	b'_M	a'_M	b'_M	$\Delta E'_{chrom.}$
Asian skin	12.92	10.16	13.04	10.09	0.14	10.29	10.86	10.29	10.36	0.51
Banana	3.88	28.29	5.07	31.77	3.68	4.44	30.83	5.08	32.59	1.88
Blueberry	−2.85	−12.81	−3.54	−12.62	0.73	−2.42	−10.54	−3.32	−12.29	1.97
Blue jeans	−7.63	−17.17	−8.11	−19.71	2.59	−7.12	−14.75	−7.79	−17.36	2.69
Broccoli	−9.62	14.38	−8.62	9.20	5.27	−8.30	16.10	-8.27	15.38	0.72
Butternut	15.75	20.64	14.73	20.47	1.04	14.19	19.70	15.38	20.57	1.47
Carrot	27.49	26.63	25.09	25.01	2.89	24.99	24.65	25.64	25.63	1.18
Caucasian skin	12.72	9.40	12.19	9.23	0.56	8.46	8.25	9.88	8.52	1.33
Concrete	−0.82	−3.54	−0.80	−3.33	0.20	−0.48	−1.35	0.08	−1.17	0.59
Green salad	−13.52	23.13	−13.59	24.19	1.06	−9.56	26.06	−10.43	26.70	1.09
Red cabbage	11.91	−16.93	12.86	−13.65	3.42	10.41	−14.65	10.56	−14.08	0.59
Red rose	37.11	13.21	36.66	12.75	0.65	34.89	11.25	34.68	11.12	0.23

Note: Additionally tabulated are the corresponding chromatic differences $\Delta E'_{chrom.}$ of the color appearance rating results due to different adaptation conditions (upper part) and the observers' cultural background (lower part).

sensitive to deviations from the objects' memory color centers compared with their German counterparts. Especially the test object of broccoli shows a disproportionally large acceptance boundary for the Chinese observers when being adapted to the 3200 K white point. This indicates that the Chinese observers might have been confronted with some difficulties in recalling the typical bluish-green broccoli object color when being adapted to the more warm-white illumination condition.

On a closer inspection of the tolerance ellipses given in Figures 4 and 5, it can further be stated that the test objects exhibiting the most narrow Gaussian similarity distributions are those of Asian and Caucasian skin. This indicates that, independent of their cultural background and the adapted white point, observers are more sensitive to changes in the appearance of skin colors than to changes in the perceived chromaticities of any of the remaining test objects. In addition, the corresponding tolerance ellipses obtained for both cultural observer groups show pronounced similarities not just in size but also in shape, orientation, and location. As a consequence, it can be concluded that people across various cultures and

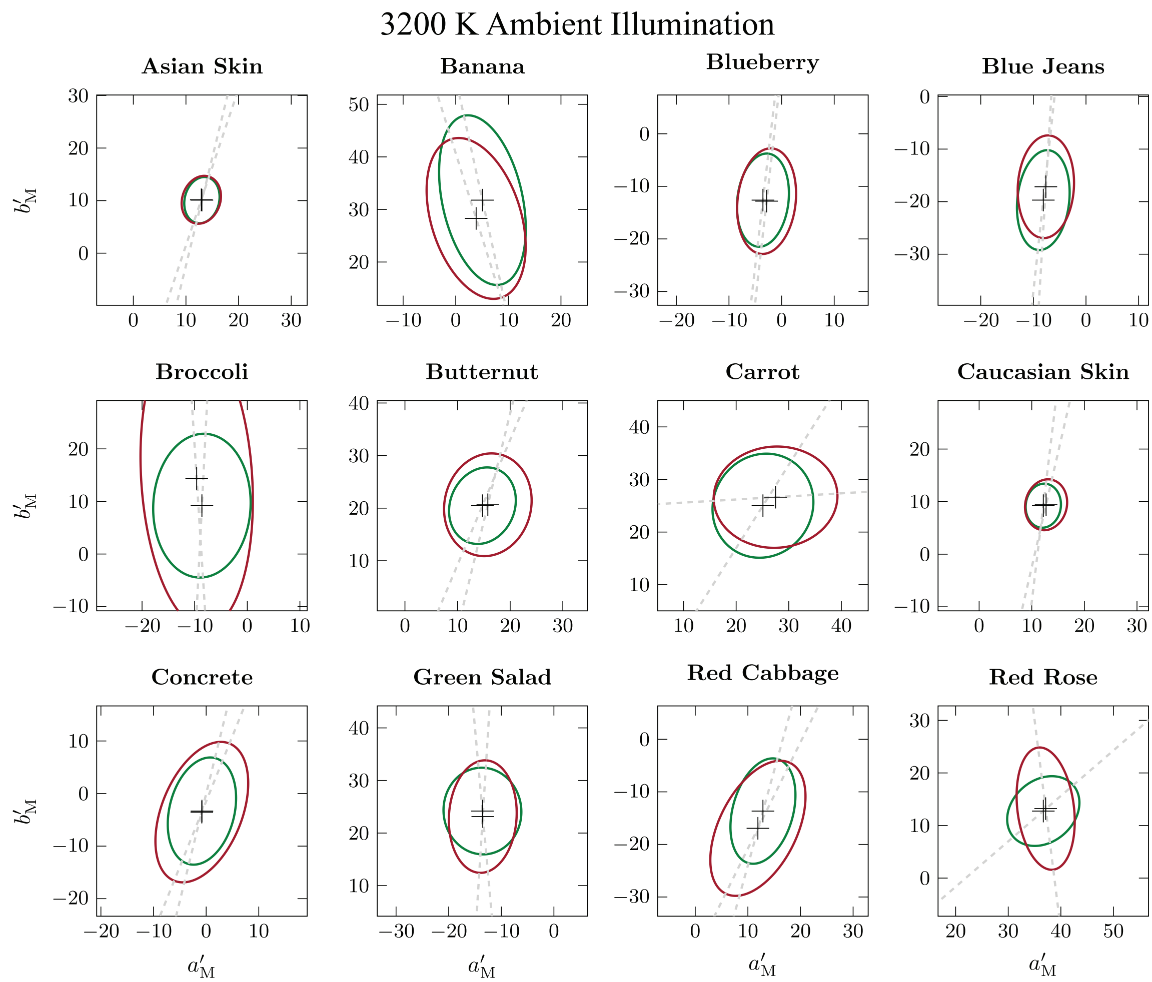

FIGURE 4 Comparison of the acceptance boundary ellipses of the 12 familiar test objects between Chinese (——) and German (——) observers at 3200 K ambient illumination. These tolerance ellipses were calculated from the fitted Gaussian distributions in such a way that their contour lines represent an average observer rating of "3," which is considered to be the just acceptable limit

nationalities seem to have a quite common and consistent notion of how skin colors should ideally look like. This basically emphasizes the importance of skin colors in the context of preference and memory, which has already been confirmed by various authors.[8,57-64]

Despite the observed color differences of the memory color centers reported in Table 2, most of the familiar test objects generally show a large overlap between the chromatic tolerance ellipses of the two different observer groups (and a somewhat smaller but still considerably large overlap between the ellipses of the two different adaptation conditions). For the test objects of Asian skin, blueberry, broccoli, butternut squash, Caucasian skin, and concrete in case of the 3200 K ambient illumination and for the test objects of broccoli, butternut squash, Caucasian skin, and concrete regarding the 5600 K case, the tolerance ellipses of the German observer group are even completely enclosed by the respective ellipses of the Chinese observer group. Based on these findings it is hardly surprising that the observed crosscultural differences expressed in terms of $\Delta E'_{\text{chrom.}}$ values between the Chinese and German memory color centers are found to be much smaller than the extent of the overlap of the corresponding acceptance boundary ellipses. With these crosscultural variations being also smaller than the differences caused by the two different adaptation conditions, it can be concluded that even though a statistical significant impact of the observers' cultural background on the color appearance ratings might be confirmed, the corresponding effect size is expected to be rather low. Note that this would be in accordance to the findings of the previous studies discussed in Section 3.3.

4.2 | Interobserver and intraobserver variability

In order to get a notion of the underlying interobserver and intraobserver variabilities being present

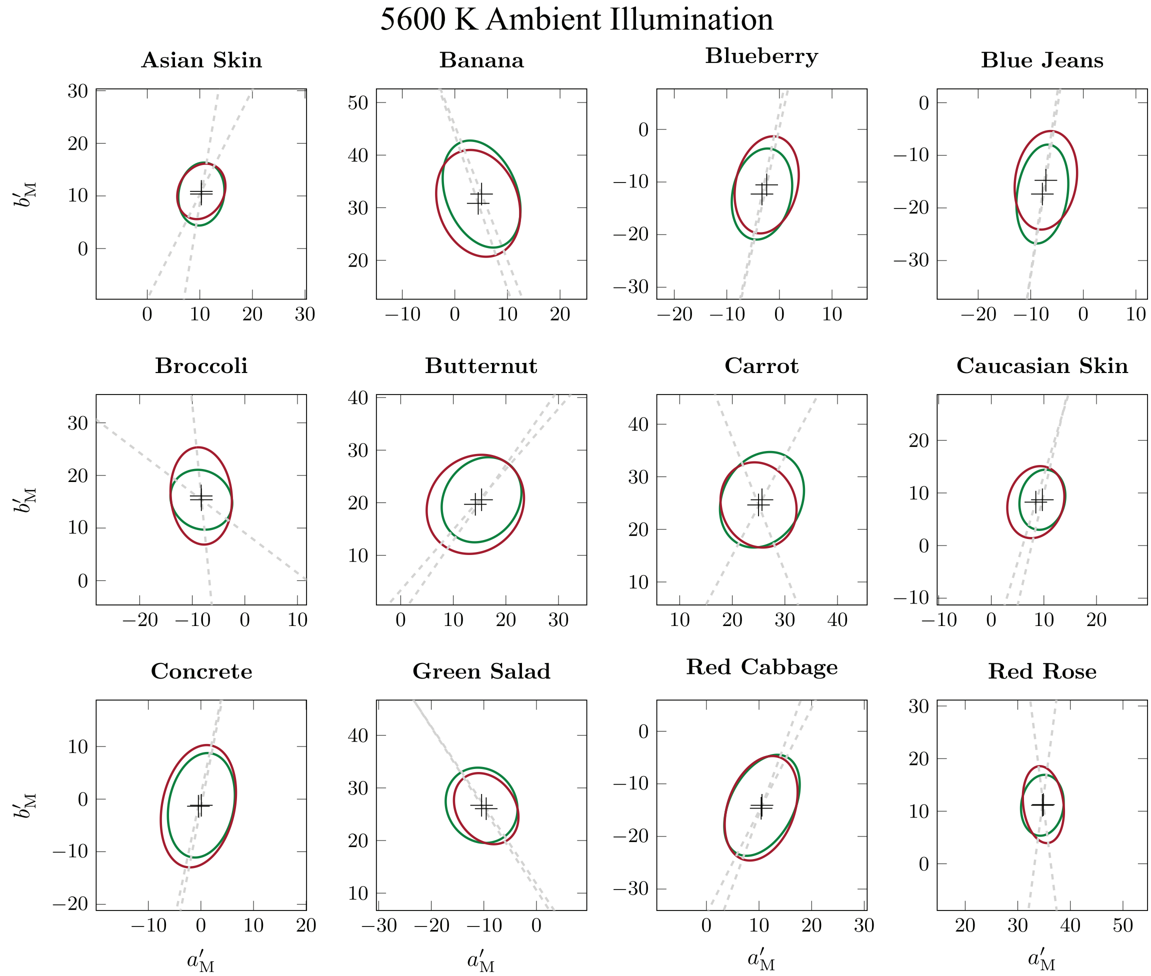

FIGURE 5 Comparison of the acceptance boundary ellipses of the 12 familiar test objects between Chinese (——) and German (——) observers at 5600 K ambient illumination. These tolerance ellipses were calculated from the fitted Gaussian distributions in such a way that their contour lines represent an average observer rating of "3," which is considered to be the just acceptable limit

in the collected rating data, both $PF/3_{inter}$ and $PF/3_{intra}$ performance factors were calculated. In both cases, a value of zero indicates perfect agreement in the memory color assessments, that is, the smaller the PF/3 values, the more consistent the ratings either between ("inter") or within ("intra") individual observers, where the latter was assessed on repeated trials. The formulae required for PF/3 calculations in the present case have recently been summarized by Babilon and Khanh[35] and are applied here in exactly the same way. In Table 3, their average values are summarized for each of the 12 familiar test objects assessed by Chinese and German observers at 3200 K and 5600 K ambient illumination.

As can be seen, both interobserver and intraobserver variabilities in the observers' color appearance ratings are quite consistent between the different cultural observer groups and adaptation conditions. While for the German observers the $PF/3_{inter}$ and $PF/3_{intra}$ values range from 28.10 to 38.12 (Ø31.98 at 3200 K, Ø34.06 at 5600 K) and from 15.18 to 22.12 (Ø17.84 at 3200 K, Ø18.34 at 5600 K), respectively, somewhat larger interobserver and intraobserver variabilities ranging from 30.41 to 43.41 (Ø36.38 at 3200 K, Ø36.68 at 5600 K) and from 17.16 to 23.05 (Ø19.11 at 3200 K, Ø19.45 at 5600 K) must be reported for the Chinese observers. However, it should be noted that statistical analysis performed on the variability distributions for each test object revealed no systematic impact of the adapted white point or the observers' cultural background on the observed rating variabilities. Only for the test object of broccoli, for which Chinese observers are supposed to have some difficulties in recalling its typical object color when being adapted to the 3200 K condition (see Section 4.1), a significantly and considerably larger (in terms of effect size) interobserver PF/3 value can be identified.

Nevertheless, it can be concluded that the observed variability in the subjects' ratings is more kind of an inherent, relatively stable feature of the corresponding experimental design and, therefore, less sensitive to the

TABLE 3 Average PF/3$_{inter}$ and PF/3$_{intra}$ performance factors calculated from the visual appearance ratings for the 12 familiar test objects assessed by Chinese and German observers at 3200 K and 5600 K ambient illumination

	Chinese observer group				German observer group			
	3200 K		5600 K		3200 K		5600 K	
Test object	PF/3$_{inter}$	PF/3$_{intra}$	PF/3$_{inter}$	PF/3$_{intra}$	PF/3$_{inter}$	PF/3$_{intra}$	PF/3$_{inter}$	PF/3$_{intra}$
Asian skin	38.15	22.49	37.70	23.05	36.30	21.36	35.61	20.15
Banana	35.39	17.31	34.67	18.78	31.72	17.31	34.45	18.35
Blueberry	39.54	20.51	43.41	20.10	35.22	18.99	35.29	18.79
Blue jeans	38.59	20.45	37.50	20.44	30.77	19.41	35.83	17.00
Broccoli	42.86	17.48	35.17	17.36	27.45	15.99	30.64	18.97
Butternut	30.41	17.16	36.95	17.75	30.17	16.40	29.49	16.60
Carrot	33.69	19.92	36.72	20.82	32.72	17.26	33.76	17.31
Caucasian skin	39.09	20.69	33.17	17.28	38.12	22.01	35.42	17.25
Concrete	39.15	18.03	36.79	17.51	31.11	15.67	34.05	17.42
Green salad	32.61	17.67	39.29	20.60	28.10	15.18	37.55	22.12
Red cabbage	34.61	19.73	34.04	20.53	33.85	18.45	34.95	18.55
Red rose	32.41	17.93	34.80	19.22	28.18	16.08	31.62	17.61
Average PF	36.38	19.11	36.68	19.45	31.98	17.84	34.06	18.34

observers' cultural background or to changes in the adaptation conditions as one might expect at first glance. Compared to color discrimination experiments,[65-68] where typical performance factors are reported to be of the order of 30 PF/3 units for the interobserver and 20 PF/3 units for the intraobserver variability, similar precision and repeatability in the observers' color appearance ratings can be stated here (see overall average values in Table 3). This is a remarkably good result considering that ratings were performed in relation to a reference kept in memory only.

Compared with the memory color rating results of Smet et al,[9] where average interobserver and intraobserver PF/3 values of 40 and 23 were reported, slightly smaller variabilities could be observed in the present experiments for both cultural observer groups and adaptation conditions. Even in the case of broccoli at 3200 K, where the Chinese observers might have had difficulties in recalling its typical object color, the within- and between-observer rating variabilities are still in an acceptable range ensuring good consistency in the observer ratings for all experimental conditions and selected test objects.

5 | CHARACTERISTICS OF MEMORY COLORS FOR DIFFERENT ADAPTATION CONDITIONS

Based on the findings reported above, the current section is dedicated to the statistical analysis of the rating data to explore further the impact of the different adaptation conditions on the fitted similarity distributions and the respective memory color centers. With the color appearance rating experiments being conducted at two different ambient illuminations with different CCTs but apart from that at equal experimental conditions for both cultural observer groups, the question is whether the observed differences between these two experimental runs reported in Section 4 can be explained by the variation of the adapted white point (see Figure 2A,B) or if they occur simply by pure chance.

For a better visualization, contour line plots can be used to compare the fitted similarity distribution functions of each test object for the two different adaptation conditions. As an example, Figure 6 shows these contour line plots for a selection of different test objects assessed by the German observers. As can be seen, slight deviations in the four ellipse dimensions shape, orientation, size, and location are observed for the test objects. In some cases though they are more pronounced than in others without really showing any kind of systematic consistencies that could be attributed to the observers' state of chromatic adaptation. Hence, when taking into account the whole set of test objects, no general trend can be derived between the results of both experimental runs, neither for the Chinese nor for the German observers. In other words, no indication is given that a certain adaptation condition would alter the ellipse dimensions of the corresponding similarity distributions of the different test objects in a general, well defined way.

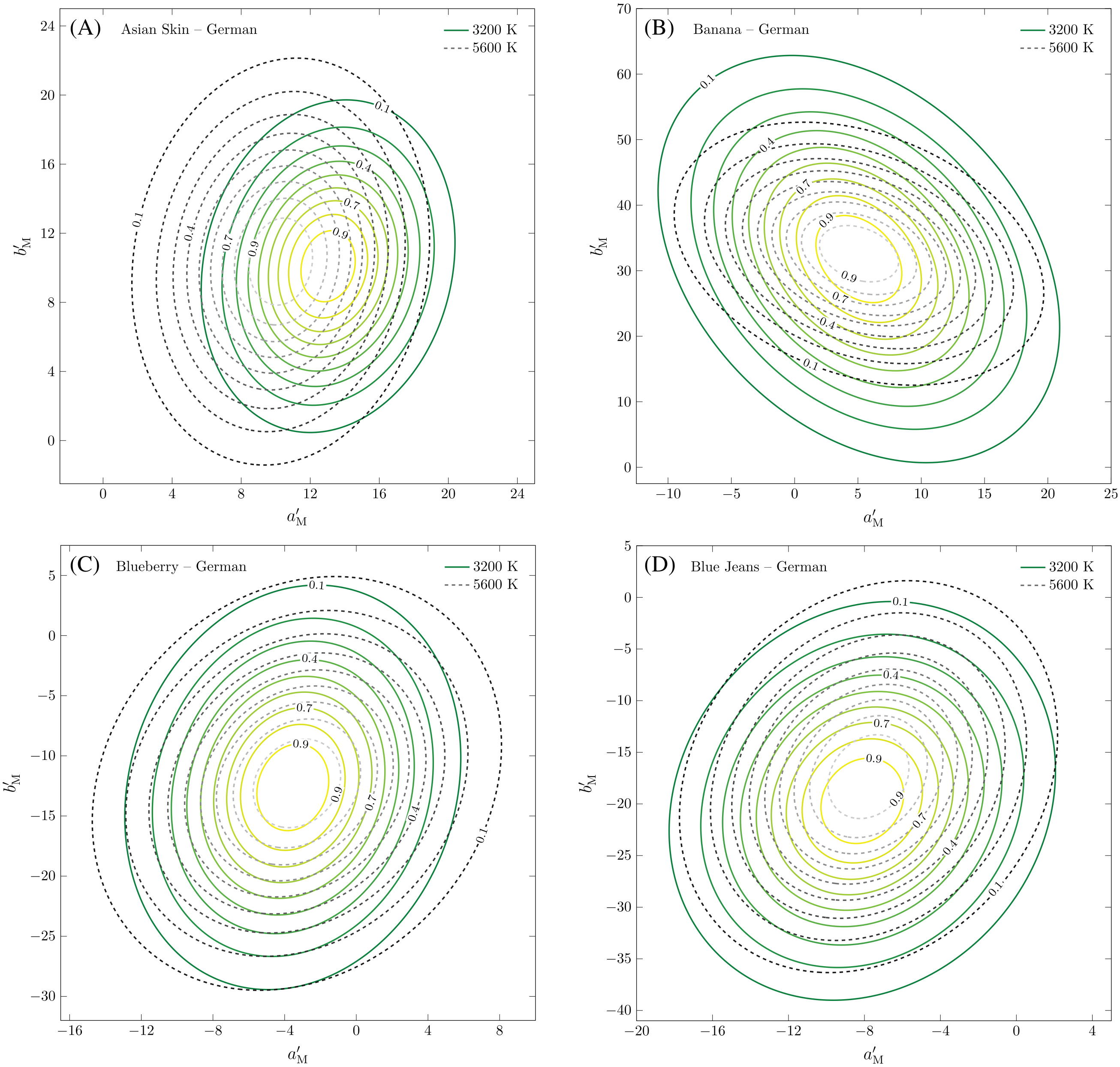

FIGURE 6 Exemplary contour line plots of the fitted, normalized similarity distribution functions obtained for the two different adaptation conditions at 3200 K (green-to-yellow colormap, solid line) and 5600 K (dark-to-light-gray colormap, dashed line) for the test objects of, A, Asian skin, B, banana, C, blueberry, and D, blue jeans as assessed by the German observers

Instead, a somewhat random pattern regarding the variations in shape, orientation, size, and location is observed. In this context, the most obvious nonconformities between the similarity distribution functions assigned to different CCTs are found for the test objects of Asian skin, banana, broccoli, carrot, Caucasian skin, green salad, and red rose, while for the remaining test objects the observed variations are less conspicuous.

Basically, these findings are in accordance to the previous work of Sanders[5] who, with the results of his memory color rating experiments, also reported nonsystematic deviations of various degree between the tolerance ellipses fitted to his selection of test objects assessed under two different reference light sources. Thus, it can be concluded that the way the preferred color appearance of certain test objects is recalled by the observers while being adapted to a specific ambient illumination is more object-dependent rather than following a universally valid scheme. This means that for some of the test objects that are used in the current work, like blueberry, blue jeans, and butternut squash, a quite good consistency in the color appearance ratings and, therefore, in the fitted similarity distributions is observed between the different adaptation conditions, while for the remaining test objects, where the manifestation of observed variations in the similarity distributions changes from object to object, no such memory color constancy could be inferred.

In order to determine whether or not these object-dependent deviations between the two adaptation conditions at 3200 K and 5600 K are significant, Box's M-test and Hotelling's T^2-test with Bonferroni correction were applied here to compare the covariances and means of the underlying multivariate data samples. Instead of testing the actual rating data, model testing of the fitted similarity distributions should be performed here. This is due to the fact that, also with regard to future applications, we are more interested in the analysis of potential deviations in the general characteristics of the memory colors derived from the similarity ratings than in the rating data itself. Given the excellent goodness-of-fit measures,[38] deviations of the actual rating data from the Gaussian modeling, however, are assumed to be quite small.

For statistical testing, a multivariate random number generator[69] was therefore adopted to calculate for each test object and adaptation condition a representative data sample based on the corresponding fitted similarity distribution. In each case, the number of samples was chosen to equal the number of chromaticity points originally presented to the observers during the rating experiments.

In Table 4 the corresponding results are summarized. In addition to the various test statistics and P values, the multivariate effect sizes are also tabulated, which are expressed in terms of the Mahalanobis distance D^2 between the respective group means. The Mahalanobis distance gives a measure for the separation of the independent group means as a distance in space in relation to the covariances of the two similarity distribution functions that should be compared. It can easily be shown[70,71] that

$$D^2 = \frac{(n_{3200} + n_{5600})t^2}{n_{3200} n_{5600}}, \tag{1}$$

where t^2 is the test statistic of Hotelling's T^2-test and n_{3200} and n_{5600} are the sample sizes assumed for the bivariate similarity distributions at 3200 K and 5600 K ambient illumination, respectively.

Let us consider the German results first. As can be seen, the application of Box's M-test revealed dealing with unequal covariances only for the test objects of broccoli and red rose giving in each case a P value smaller than .001.

TABLE 4 Resulting P values and test statistics of Box's M-test and Hotelling's T^2-test applied to the similarity distribution functions of the 12 familiar test objects assessed by Chinese and German observers under two different adaptation conditions at 3200 K and 5600 K ambient illumination to check for adaptation-dependent differences

	Chinese observer group					**German observer group**				
	Box's M-test		**Hotelling's T^2-test**			**Box's M-test**		**Hotelling's T^2-test**		
	Test statistic χ^2	**P value**	**Test statistic t^2**	**P value**	**Effect size D^2**	**Test statistic χ^2**	**P value**	**Test statistic t^2**	**P value**	**Effect size D^2**
Asian skin	1.209	.751	**13.149**	**.002**	**0.405**	4.261	.235	**12.417**	**.002**	**0.382**
Banana	5.152	.161	2.736	.255	0.084	11.857	.008	0.383	.826	0.012
Blueberry	1.023	.796	1.872	.392	0.058	2.456	.483	0.210	.900	0.006
Blue jeans	1.210	.751	1.977	.372	0.061	0.205	.977	1.874	.391	0.058
Broccoli	**115.634**	**<.0001**	0.588[a]	.745[a]	0.018[a]	**65.095**	**<.0001**	10.566[a]	.005[a]	0.325[a]
Butternut	0.707	.872	0.788	.674	0.024	1.327	.723	0.385	.825	0.012
Carrot	11.242	.011	4.914	.086	0.151	3.228	.358	0.253	.881	0.008
Caucasian	9.426	.024	**21.554**	**<.0001**	**0.663**	5.425	.143	8.920	.012	0.274
Concrete	5.590	.133	0.683	.711	0.021	0.360	.948	1.592	.451	0.049
Green salad	**13.916**	**.003**	**12.801[a]**	**.002[a]**	**0.394[a]**	2.240	.524	8.140	.017	0.250
Red cabbage	6.309	.098	4.295	.117	0.132	2.639	.451	2.965	.227	0.091
Red rose	12.389	.006	8.883	.012	0.273	**18.069**	**0.001**	6.310[a]	.043[a]	0.194[a]

Note: If not indicated otherwise (by a footnote) the standard, homoscedastic version of Hotelling's T^2-test was used. Bold table entries indicate significant differences between both adaptation conditions at a Bonferroni corrected significance level of $\alpha = .004$. For each test object, the corresponding effect size is given in terms of the Mahalanobis distance D^2.

[a] Heteroscedastic version of Hotelling's T^2-test must be applied here.

While for the latter this deviation in its corresponding covariance matrix is mainly due to differences in size and orientation of the respective similarity distribution function, for the former also a change in shape can be observed.

Regarding the outcome of Hotelling's T^2-test, the null hypothesis of equal sample mean vectors and, therefore, of equal memory color centers must be rejected only for the test object of Asian skin. In this case, the corresponding color difference reported in Section 4 between the memory color centers obtained for the two different adaptation conditions is considered to be statistically significant showing a medium large effect size. For the remaining test objects no such significance in the mean vectors of their similarity distributions can be concluded. Thus, with only three out of 12 familiar test objects showing either significantly different covariance matrices or significantly different locations of the memory color centers, a noteworthy and systematic impact of the different adaptation conditions on the observers' memory color assessments cannot be confirmed for the German observers, at least not for the CCT range considered here.

Similar conclusions hold for the Chinese assessments, where significantly unequal covariances can only be reported for the test objects of broccoli and green salad. While for the latter the respective similarity distribution functions obtained for the two distinct adaptation conditions mainly differ in size and orientation, an additional change in shape is observed for the test object of broccoli. Regarding the differences of the sample mean vectors, significant deviations in the Chinese assessments can be reported for the test objects of Asian skin, Caucasian skin, and green salad. However, with in total only four out 12 test objects showing significance when comparing the results at 3200 K and 5600 K, respectively, an overall impact of these two different adaptation conditions on the observers' memory color assessments again remains questionable.

What is kind of interesting though is that for most test objects, with the exception of Asian skin, blueberry, blue jeans, and butternut, the sizes of the corresponding similarity distribution functions fitted to the Chinese ratings at 5600 K ambient illumination are across-the-board considerably smaller than their 3200 K counterparts indicating that, in contrast to the German assessments where no such tendency is observed, Chinese people generally are less sensitive to deviations from the respective color centers when being adapted to the more warm-white illumination condition. Even though no clear statistical effect could be confirmed so far, this might indicate that a certain tendency exists in the rating data that, however, cannot be represented statistically when comparing the fitted similarity distributions directly with each other.

In such cases, the application of an extra sum-of-squares F-test can be beneficial. This test has already been applied successfully for the treatment of similar problems.[30] Basically, it can be used to evaluate the goodness-of-fit of two alternative, nested models fitting the same data, where one model is a simpler version (fewer parameters) of the other (more parameters). Based on statistical hypothesis testing, the extra sum-of-squares F-test compares the relative improvement in the sum-of-squares (SS) of the more complicated model with the relative loss of degrees of freedom that goes hand in hand. If the simpler model is correct (null hypothesis), the amount of improvement in the SS of the more complex model (alternative hypothesis) is assumed to be observed merely by chance, which in turn is determined by the degrees of freedom in each model. In other words, the extra sum-of-squares F-test compares the difference in the SS between booth models with the difference one would expect by chance, which can mathematically be expressed by

$$F = \frac{(\mathrm{SS}_{\mathrm{null}} - \mathrm{SS}_{\mathrm{alt.}})/\mathrm{SS}_{\mathrm{alt.}}}{(df_{\mathrm{null}} - df_{\mathrm{alt.}})/df_{\mathrm{alt.}}} = \left(\frac{\mathrm{SS}_{\mathrm{null}}}{\mathrm{SS}_{\mathrm{alt.}}} - 1\right)\left(\frac{df_{\mathrm{null}}}{df_{\mathrm{alt.}}} - 1\right)^{-1}, \tag{2}$$

where $\mathrm{SS}_{\mathrm{null}}$ and $\mathrm{SS}_{\mathrm{alt.}}$ are the residual sum-of-squares between the data to be fit and the respective model, while df_{null} and $df_{\mathrm{alt.}}$ denote the corresponding degrees of freedom. Hence, the test statistic F equals the relative difference in the sum-of-squares between the simple and the more complex model divided by their relative difference in degrees of freedom.

In the present case of the color appearance rating experiments conducted at different CCTs of the ambient illumination, the simple model assumes that for each test object the variance of the entire rating data set obtained by pooling the respective results of both adaptation conditions can be explained by a single multivariate Gaussian function defined by seven fit parameters (cf Babilon et al[35,72]). The more complex model, on the other hand, postulates a separate Gaussian for each adaptation condition to explain the total variance in the rating data leading to a total number of 14 fit parameters, that is, seven for each CCT. Based on these considerations, the null hypothesis assuming the correctness of the simple model can be evaluated for each familiar test object and cultural observer group.

The corresponding results are summarized in Table 5 indicating a statistically significant effect for all 12 test objects and both cultural observer groups. In addition, the effect size $\eta^2 = (\mathrm{SS}_{\mathrm{null}} - \mathrm{SS}_{\mathrm{alt.}})/\mathrm{SS}_{\mathrm{null}}$ is calculated and ranges between 0.049 and 0.382 with an average value of 0.212 for the Chinese and between 0.089 and 0.294 with an average value of 0.181 for the German observers, which, following Cohen's rule of thumb,[73] is concluded to represent in all cases a medium large statistical effect. Hence, evidence is given that, despite no significance was

TABLE 5 Results of the extra sum-of-squares *F*-test for the effect of the adapted white point/ambient illumination

	Chinese observer group			German observer group		
	Test statistic *F*	*P* value	Effect size η^2	Test statistic *F*	*P* value	Effect size η^2
Asian skin	1.525	.0108	0.049	3.727	<.0001	0.162
Banana	6.307	<.0001	0.247	3.716	<.0001	0.163
Blueberry	11.925	<.0001	0.382	8.034	<.0001	0.294
Blue jeans	5.638	<.0001	0.224	6.921	<.0001	0.262
Broccoli	1.988	.0002	0.090	6.872	<.0001	0.256
Butternut squash	4.525	<.0001	0.190	2.185	<.0001	0.102
Carrot	10.057	<.0001	0.342	3.248	<.0001	0.144
Caucasian skin	3.103	<.0001	0.140	3.675	<.0001	0.162
Concrete flowerpot	8.337	<.0001	0.259	6.841	<.0001	0.258
Green salad	3.572	<.0001	0.158	1.857	.0005	0.089
Red cabbage	6.478	<.0001	0.251	2.660	<.0001	0.121
Red rose	5.253	<.0001	0.214	3.629	<.0001	0.158

Note: Statistical significance is given for all 12 test objects. In addition, the corresponding effect size η^2 is tabulated indicating a medium large statistical effect.

found in the direct comparison of the fitted similarity distributions, the impact of different adaptation conditions on memory color assessments cannot be precluded entirely. Nonetheless, it is most likely of no practical relevance, at least for the CCT range considered in this work.

6 | IMPACT OF THE CULTURAL BACKGROUND

After having discussed the impact of the adapted white point on the color appearance ratings of familiar test objects, a similar analysis should be performed in the following regarding the influence of the observers' cultural background. With the color appearance rating experiments being conducted under more or less identical experimental conditions for both cultural observer groups, the differences observed between the Chinese and German subjects reported in Section 4 give indication that a cultural component in the assessment of memory colors may exist. Even though these differences are quite small and, therefore, as stated by Smet et al[30] are likely to be of no practical importance, statistical analysis should be provided in order to get an idea about the significance and the size of the effect. Again Box's *M*-test and Hotelling's T^2-test were applied to compare the covariances and mean vectors of the bivariate Gaussian similarity distributions that were fitted to the Chinese and German color appearance ratings gathered for both adaptation conditions.

In Table 6, the corresponding results are summarized. As expected from the discussions provided in Section 4, the null hypothesis of equal sample mean vectors cannot be rejected for any of the familiar test objects assessed under both adaptation conditions. Thus, no significant differences in the memory color centers reported in the lower part of Table 2 can be found between Chinese and German observers. This is further emphasized by the calculated effect sizes which in all cases are quite small ($D^2 < 0.09$) indicating that for each test object the overlap of the fitted similarity distributions of both cultural observer groups is too large to be able to resolve the small difference between their mean vectors. Regarding the size, shape, and orientation of the corresponding similarity distribution functions given by their covariance matrices, significant differences between Chinese and German observers are only obtained for the test objects of broccoli, concrete, and red rose at 3200 K ambient illumination.

Obviously, these results are in accordance to the findings of Figures 4 and 5, where, due to the relatively large overlap between the Chinese and German acceptance boundary ellipses, deviations are only observed for the ellipse parameters of shape, size, and orientation but not for their locations. For all test objects, the observed cross-cultural differences expressed in terms of $\Delta E'_{\text{chrom.}}$ values between the respective memory color centers were generally found to be much smaller than the extent of the overlap of the corresponding acceptance boundary ellipses and/or similarity distributions.

Hence, the original claim of Smet et al[30] that the impact of the cultural background of the observers on memory color assessments is negligibly small from a practical point of view seems to be verified also for this new set of data which, for the first time, takes into

TABLE 6 Resulting *P* values and test statistics of Box's *M*-test and Hotelling's T^2-test applied to the similarity distribution functions of the 12 familiar test objects assessed under two different adaptation conditions at 3200 K and 5600 K ambient illumination to check for cultural-dependent differences between the average color appearance ratings of Chinese and German observers

	3200 K adapted white point					5600 K adapted white point				
	Box's *M*-test		Hotelling's T^2-test			Box's *M*-test		Hotelling's T^2-test		
	Test statistic χ^2	*P* value	Test statistic t^2	*P* value	Effect size D^2	Test statistic χ^2	*P* value	Test statistic t^2	*P* value	Effect size D^2
Asian skin	0.231	.973	0.023	.988	0.001	1.708	.635	0.358	.836	0.011
Banana	1.849	.604	2.839	.242	0.087	0.938	.816	0.943	.624	0.029
Blueberry	2.238	.525	0.882	.643	0.027	0.088	.993	1.400	.497	0.043
Blue jeans	1.488	.685	1.674	.433	0.052	4.229	.238	2.128	.345	0.065
Broccoli	**36.050**	**<.0001**	0.875[a]	.646[a]	0.027[a]	11.389	.010	0.248	.884	0.008
Butternut	8.618	.035	0.967	.617	0.030	3.120	.374	0.487	.784	0.015
Carrot	2.065	.559	2.369	.306	0.073	4.210	.240	1.075	.584	0.033
Caucasian	0.276	.965	0.631	.730	0.019	1.580	.664	1.494	.474	0.046
Concrete	**13.591**	**.0035**	0.149[a]	.928[a]	0.005[a]	4.181	.243	0.096	.953	0.003
Green salad	6.208	.102	0.316	.854	0.010	0.712	.870	0.925	.630	0.028
Red cabbage	12.987	.005	1.725	.422	0.053	0.533	.912	0.163	.922	0.005
Red rose	**24.685**	**<.0001**	0.100[a]	.951[a]	0.003[a]	6.636	.085	0.144	.931	0.004

Note: If not indicated otherwise (by a footnote) the standard, homoscedastic version of Hotelling's T^2-test was used. Bold table entries indicate significant differences between both cultural observer groups at a Bonferroni-corrected significance level of $\alpha = .004$. The corresponding effect size is given terms of the Mahalanobis distance D^2.

[a] Heteroscedastic version of Hotelling's T^2-test must be applied here.

account realistic viewing and adaptation conditions. Additional confirmation is therefore supposed to be provided by again applying the extra sum-of-squares *F*-test.

For this purpose, following the procedure of Smet et al,[30] a global set of rating data is defined for each test object and adaptation condition by pooling the corresponding mean ratings of the Chinese and German observers. The simple model now assumes that the observed variance of this global set of rating data can be explained by a single bivariate Gaussian function defined by seven fit parameters as described previously. The more complex model, on the other hand, postulates a separate Gaussian for each observer group to explain the total variance in the combined set of rating data leading to a total number of 14 fit parameters. The null hypothesis assuming the correctness of the simple model was evaluated and the corresponding results are summarized in Table 7.

As can be seen, in most cases the null hypothesis must be rejected which indicates that the total variance in the rating data of the pooled Chinese and German observers cannot be described adequately by the simple model for all test objects. Hence, it can be concluded that with the exceptions of butternut squash, carrot, and red rose at 3200 K and of Caucasian skin at 5600 K ambient illumination the extra sum-of-squares *F*-test reveals a nonnegligible, statistically significant intercultural effect on the color appearance ratings between Chinese and German observers. This effect mainly manifests in deviations of size, shape, and orientation of the respective sample distributions rather than in shifts of their centroid locations. However, the corresponding effect sizes in terms of η^2 are quite small. On average, the effect size is approximately 44% smaller than the effect size corresponding to the impact of the adapted white point (see Table 5). This gives an intrinsic hierarchy for the importance of these two different effects on the color appearance ratings of the familiar test objects considered in the current work.

Even though intercultural variations are of minor importance compared to the influence of the chromatic adaptation conditions, both effects are found to be statistically significant from the analysis of the extra sum-of-squares *F*-test applied to the rating data.

TABLE 7 Results of the extra sum-of-squares *F*-test for the effect of the cultural background on the color appearance ratings of familiar real-test objects

	3200 K adapted white point			5600 K adapted white point		
	Test statistic *F*	*P* value	Effect size η^2	Test statistic *F*	*P* value	Effect size η^2
Asian skin	4.709	<.0001	0.196	4.825	<.0001	0.200
Banana	3.199	<.0001	0.144	2.004	<.0001	0.094
Blueberry	2.363	<.0001	0.109	2.159	<0.0001	0.098
Blue jeans	2.989	<.0001	0.132	1.761	.0011	0.083
Broccoli	1.941	<.0001	0.087	2.566	<0.0001	0.115
Butternut squash	0.670	.985	0.034	1.422	.027	0.067
Carrot	0.374	.998	0.019	2.418	<0.0001	0.112
Caucasian skin	4.849	<.0001	0.201	0.918	.678	0.047
Concrete flowerpot	1.870	.0003	0.086	2.053	<0.0001	0.094
Green salad	2.397	<.0001	0.112	2.212	.0005	0.104
Red cabbage	1.923	.0002	0.112	3.787	<0.0001	0.164
Red rose	0.935	.643	0.158	1.434	.0256	0.069

Note: Statistical significance is given for all 12 test objects. In addition, the corresponding effect size η^2 is tabulated indicating a small to medium large statistical effect.

However, when taking into account the corresponding effect sizes as well as the results of the statistical testing directly performed on the fitted similarity distributions, further indication is given that the impact of the observers' cultural background is likely to be of no practical relevance in the context of memory color assessments, confirming the conclusions drawn by Smet et al[30,32] in the context of realistic viewing and adaptation conditions.

7 | CONCLUSION AND OUTLOOK

In this article, the impact of both the adapted white point and the observers' cultural background on the color appearance ratings of a set of 12 different familiar test objects assessed under realistic viewing and adaptation conditions has been investigated. The corresponding experiments were performed at two different ambient illuminations with CCTs of 3200 K and 5600 K, respectively, where each test object's color appearance was modulated by an LCD projector and rated separately by a group of Chinese and German observers. By providing a comprehensive statistical analysis of the collected rating data obtained for the two different cultural observer groups and adaptation conditions, it could be shown that the observed intercultural variations between Chinese and German subjects as well as the impact of the adaptation conditions on the color appearance ratings were both found to be significant. With the latter showing an approximately 44% larger effect size than the former, an intrinsic hierarchy for the importance of these two different impact factors can be deduced.

However, when taking into account the absolute effect size values as well as the results of the statistical testing performed directly on the fitted similarity distributions, indication is given that both effects are likely to be of no practical relevance in the context of memory color assessments. Regarding the impact of the observers' cultural background, this basically confirms the conclusions drawn by Smet et al[30,32] also for realistic viewing and adaptation conditions. For the impact of the adapted white point, on the other hand, this holds only true for the CCT range considered in this work. As one proceeds to significantly higher or lower CCTs, incomplete chromatic adaptation might have a larger impact also in the case of an immersive viewing scenario such as the one considered here.

As shown recently in several series of viewing-booth-like experiments,[33,34,74,75] the degree of adaptation and, therefore, perceived color appearance highly depends on the illumination chromaticity. A similar dependency can be expected for memory color assessments performed in realistic viewing situations and, consequently, requires a systematic study of the impact of (incomplete) chromatic adaptation in these situations at different illumination conditions.

Hence, an extension of the reported experiments to a greater number of different adaptation conditions would be preferable. Profound knowledge of the underlying dependencies is required to develop—as a long-term goal—an improved version of a memory-based color quality metric. A first attempt of defining such a metric based on the results obtained for the two adapting field chromaticities considered here was presented elsewhere.[72,76] However, it should be stressed that the use of only these two adaptation conditions is certainly not enough for covering the huge variety of different lighting situations one is confronted with in real-world applications. In addition, the metric proposal still shows several weaknesses, such as overfitting, discontinuity at 4000 K, incomprehensible object weighting (pure mathematical optimization has been performed so far), and so on, that still demand further research and significant adjustments/improvements before being applicable in practice.

What has been neglected so far in the current work but should definitely be considered for obvious reasons on future occasions, is the impact of different light levels in the context of memory color assessments under realistic viewing and adaptation conditions. With international standards on indoor lighting generally recommending lower illuminances to be sufficient to fulfill the visual task than those adopted here, the question arises how these different levels of illuminance may influence the assessment of memory colors. Based on the findings of previous user preference studies, a light-level-dependent shift in the test objects' typical memory color representations is expected, also with regard to the Hunt effect, and should therefore be in the focus of future research intentions.

Apart from that, further discussions on the proper test object selection with regard to a later use in a global memory color rendition measure might be beneficial. In this context, the results of the German survey reported in Section 3.1 should be combined with those of further surveys conducted among Chinese or other cultural observer groups only. Even though no relevant crosscultural differences could be reported for the current test object selection between Chinese and German observers, it would certainly be better for a universally valid color quality metric to have a test object selection that is equally representative (not just familiar) for all potential observer groups.

Finally, from the discussion given in the introduction it would be worth examining in a future report the potential bias introduced by a specific context compared to experiments on memory colors that are solely based on contextless object presentation procedures. Such a comparison would be beneficial for explaining in which manner the context of a viewing environment impacts the memory color object perception and should be performed by directly assessing memory color appearance ratings of familiar objects presented to the observers with and without context in otherwise identical conditions. Based on the outcome of such considerations, it should further be tested in a subsequent real-world lighting preference experiment which of these two approaches would eventually lead to a memory-based color quality metric that correlates best with observer preference ratings.

ORCID

Sebastian Babilon https://orcid.org/0000-0001-8002-1220

REFERENCES

[1] Smet K, Hanselaer P. Memory and preferred colours and the colour rendition of white light sources. *Light Res Technol*. 2016;48(4):393-411.
[2] Liu Q, Huang Z, Xiao K, Pointer MR, Westland S, Luo MR. Gamut volume index: a color preference metric based on meta-analysis and optimized color samples. *Opt Express*. 2017;25(14):16378-16391.
[3] Smet K, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. Correlation between colour quality metric predictions and visual appreciation of light sources. *Opt Express*. 2011;19(9):8151-8166.
[4] Newhall SM, Burnham RW, Clark JR. Comparison of successive with simultaneous color matching. *J Opt Soc Am*. 1957;47(1):43.
[5] Sanders CL. Color preferences for natural objects. *Illum Eng*. 1959;54:452.
[6] Sanders CL. Assessment of color rendition under an illuminant using color tolerances for natural objects. *Illum Eng*. 1959;54:640.
[7] Bartleson CJ. Memory colors of familiar objects. *J Opt Soc Am*. 1960;50(1):73.
[8] Bartleson CJ, Bray CP. On the preferred reproduction of flesh, blue-sky, and green-grass colors. *Photogr Sci Eng*. 1962;6(1):19.
[9] Smet K, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. Colour appearance rating of familiar real objects. *Color Res Appl*. 2011b;36(3):192-200.
[10] Judd DB. A flattery index for artificial illuminants. *Illum Eng*. 1967;62:593.
[11] Thornton WA. A validation of the color-preference index. *J Illum Eng Soc*. 1974;4(1):48.
[12] Smet KAG, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. Memory colours and colour quality evaluation of conventional and solid-state lamps. *Opt Express*. 2010;18(25):26229-26244.
[13] Smet K, Jost-Boissard S, Ryckaert WR, Deconinck G, Hanselaer P. Validation of a colour rendering index based on memory colours. Paper presented at: Proceedings of the CIE 2010 Conference: Lighting Quality and Energy Efficiency. Vienna, Austria: International Commission on Illumination CIE; 2010:136-142.
[14] Smet KAG, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. A memory colour quality metric for white light sources. *Energ Buildings*. 2012;49:216.
[15] Yendrikhovskij SN, Blommaert FJJ, de Ridder H. Representation of memory prototype for an object color. *Color Res Appl*. 1999;24(6):393-410.
[16] Ling Y, Hurlbert AC. Color and size interactions in a real 3D object similarity task. *J Vis*. 2004;4(9):721-734.
[17] Jennes JW, Shevell SK. Color appearance with sparse chromatic context. *Vis Res*. 1995;35(6):797-805.

[18] Webster MA, Mollon JD. Adaptation and the color statistics of natural images. *Vis Res*. 1997;37(23):3283-3298.
[19] Juricevic I, Webster MA. Normal variations in color vision V: simulations of adaptation to natural color environments. *Vis Neurosci*. 2009;26(1):133-145.
[20] Lin Y, Wei M, Smet KAG, Tsukitani A, Bodrogi P, Khanh TQ. Colour preference varies with lighting application. *Light Res Technol*. 2017;48(3):316-328.
[21] Davis W, Ohno Y. Toward an improved color rendering metric. Paper presented at: Proceedings of the SPIE 5941, 5th International Conference on Solid State Lighting; 2005: 59411G–1–59411G–8; Bellingham, WA, International Society for Optics and Photonics.
[22] Davis W, Ohno Y. Color quality scale. *Opt Eng*. 2010;49(3): 033602.
[23] Smith T, Guild J. The C.I.E. colorimetric standards and their use. *Trans Opt Soc*. 1931;33(3):73-134.
[24] Judd DB. Hue saturation and lightness of surface colors with chromatic illumination. *J Opt Soc Am*. 1940;30(1):2-32.
[25] Ebner F. Derivation and modelling hue uniformity and development of the IPT color space (Ph.D. thesis). Rochester Institute of Technology; 1998.
[26] Ebner F, Fairchild MD. Development and testing of a color space (IPT) with improved hue uniformity. Paper presented at: Proceedings of the 6th Color and Imaging Conference: Color Science, Systems and Applications; 1998:8-13; Scottsdale, AZ, Society for Imaging Science and Technology (IS & T).
[27] Jin H, Zhao X, Liu H. Testing of the uniformity of color appearance space. Paper presented at: Proceedings of the 2009 World Congress on Computer Science and Information Engineering; 2009:307-311; Los Alamitos, CA, IEEE Computer Society.
[28] Luo MR, Cui G, Li C. Uniform colour spaces based on CIECAM02 colour appearance model. *Color Res Appl*. 2006;31 (4):320-330.
[29] Tarczali T, Du Park S, Bodrogi P, Kim CY. Long-term memory colors of Korean and Hungarian observers. *Color Res Appl*. 2006;31(3):176.
[30] Smet KAG, Lin Y, Nagy BV, et al. Cross-cultural variation of memory colors of familiar objects. *Opt Express*. 2014;22(26): 32308-32328.
[31] Fernandez SR, Fairchild MD, Braun K. Analysis of observer and cultural variability while generating "Preferred" color reproductions of pictorial images. *J Imag Sci Technol*. 2005;49 (1):96-104.
[32] Smet KAG, Hanselaer P. Impact of cross-regional differences on color rendition evaluation of white light sources. *Opt Express*. 2015;23(23):30216-30226.
[33] Smet KAG, Zhai Q, Luo MR, Hanselaer P. Study of chromatic adaptation using memory color matches, Part I: neutral illuminants. *Opt Express*. 2017;25(7):7732-7748.
[34] Smet KAG, Zhai Q, Luo MR, Hanselaer P. Study of chromatic adaptation using memory color matches, Part II: colored illuminants. *Opt Express*. 2017;25(7):8350-8365.
[35] Babilon S, Khanh TQ. Color appearance rating of familiar real objects under immersive viewing conditions. *Color Res Appl*. 2018;43(4):551-568.
[36] Derefeldt G, Swartling T, Berggrund U, Bodrogi P. Cognitive color. *Color Res Appl*. 2004;29(1):7-19.
[37] Werner A. Spatial and temporal aspects of chromatic adaptation and their functional significance for colour constancy. *Vis Res*. 2014;104:80-89.
[38] Babilon S. Report of the results of the online survey on memory colors; 2018. https://www.lichttechnik.tu-darmstadt.de/forschung_lt/daten_zu_wissenschaftl___veroeffentlichungen/inhalt_mit_marginalienspalte_71.de.jsp. Accessed March 10, 2020.
[39] Box HC. *Set Lighting Technician's Handbook: Film Lighting Equipment, Practice, and Electrical Distribution*. Burlington, MA: Taylor & Francis; 2010.
[40] Boyce PR. The influence of illumination level on prolonged work performance. *Light Res Technol*. 1970;2(2):74-94.
[41] Balder JJ. Erwünschte leuchtdichten in büroräumen (preferred luminance in offices). *Lichttechnik*. 1957;9(9):455-461.
[42] Muck E, Bodmann HW. Die Bedeutung des Beleuchtungsniveaus bei praktischer Sehtätigkeit. *Lichttechnik*. 1961;13(10):502-507.
[43] Bodmann HW, Söllner G, Voit E. Bewertung von beleuchtungsanlagen bei verschiedenen Lichtarten. Paper presented at: Proceedings of the CIE 1963 Conference: Light and Lighting Conference with Special Emphasis on LEDs and Soloid State Lighting; 1963:502-509; Vienna, Austria, International Commission on Illumination CIE.
[44] Bodmann HW. Quality of interior lighting based on luminance. *Transc Illumin Eng Soc*. 1967;32(1):22-40.
[45] Riemenschneider W. Beleuchtungsstärken für Arbeiträume. *Bulletin des Verbands Schweizerischer Elektrizitätsunternehmen*. 1967;58(1):19-25.
[46] Westhoff JM, Horemann HW. Contribution to Bodmann. *Lichttechnik*. 1963;14:24.
[47] Khanh TQ, Bodrogi P. Nutzerpräferenzen in der heutigen Innenraumbeleuchtung. *Licht*. 2018;70(8):64-71.
[48] Witzel C, Valkova H, Hansen T, Gegenfurtner KR. Object knowledge modulates colour appearance. *i-Perception*. 2011;2 (1):13-49.
[49] Smet KAG, Deconinck G, Hanselaer P. Chromaticity of unique white in object mode. *Opt Express*. 2014;22(21):25830-25841.
[50] Smet KAG, Deconinck G, Hanselaer P. Chromaticity of unique white in illumination mode. *Opt Express*. 2015;23(10):12488-12495.
[51] Smet KAG, Zhai Q, Luo MR, Hanselaer P. Investigating chromatic adaptation using memory colours. Paper presented at: Proceedings of the 4th CIE Symposium on Colour and Visual Appearance; 2016:16-24; Vienna, Austria, International Commission on Illumination CIE.
[52] Kraft JM, Brainard DH. Mechanisms of color constancy under nearly natural viewing. *Proc Natl Acad Sci*. 1999;96(1):307-312.
[53] Likert R. A technique for the measurement of attitudes. *Arch Psychol*. 1932;22(140):5-55.
[54] Ishihara S. *Ishihara's Tests for Colour Deficiency: 38 Plates Edition*. Tokyo: Kanehara Trading Inc; 2016.
[55] Ichikawa H, Hukami K, Tanabe S. *Standard Pseudoisochromatic Plates Part II: For Acquired Color Vision Defects*. Tokyo, New York, NY: Igaku-Shoin; 1983.
[56] Linksz A. The Farnsworth panel D-15 test. *Am J Ophthalmol*. 1966;62(1):27-37.
[57] Yano T, Hashimoto K. Preference index for Japanese complexion under illuminations. *Color Res Appl*. 2016;41(2):143-153.
[58] Jost-Boissard S, Avouac P, Fontoynont M. Preferred color rendition of skin under LED sources. *J Illum Eng Soc*. 2016;12(1):79-93.

[59] He J, Lin Y, Yano T, Noguchi H, Yamaguchi S, Matsubayashi Y. Preference for appearance of Chinese complexion under different lighting. *Light Res Technol*. 2017;49(2):228-242.
[60] Yano T, Hashimoto K. Preference index for Japanese complexion color under illumination. *J Illuminat Eng Inst Japan*. 1998; 82(11):895-901.
[61] Zeng H, Luo MR. Preferred skin color enhancement of digital photographic images. *Int J Image Process*. 2013;7(4): 314-329.
[62] Zeng H, Luo MR. Colour and tolerance of preferred skin colours on digital photographic images. *Color Res Appl*. 2013;38 (1):30-45.
[63] Park DS, Kwak Y, Ok H, Kim CY. Preferred skin color reproduction on the display. *J Electron Imag*. 2006;15(4):041203.
[64] Bartleson CJ. Some observations on the reproduction of flesh colors. *Photogr Sci Eng*. 1959;3(3):114.
[65] Guan SS, Luo MR. A colour-difference formula for assessing large colour differences. *Color Res Appl*. 1999;24(5):344-355.
[66] Guan SS, Luo MR. Investigation of parametric effects using small colour differences. *Color Res Appl*. 1999;24(5):331-343.
[67] Guan SS, Luo MR. Investigation of parametric effects using large colour differences. *Color Res Appl*. 1999;24(5):356-368.
[68] Xu H, Yaguchi H, Shioiri S. Estimation of color-difference formulae at color discrimination threshold using CRT-generated stimuli. *Opt Rev*. 2001;8(2):142-147.
[69] Hernádvölgyi IT. Generating random numbers from the multivariate normal distribution; 1998. ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia013_2s09/material_de_apoio/gen_rand_multivar.pdf. Accessed March 10, 2020.
[70] Sapp M, Obiakor F, Gregas AJ, Scholze S. Mahalanobis distance: a multivariate measure of effect in hypnosis research. *Sleep Hypn*. 2007;9(2):67-70.
[71] Stevens JP. *Applied Multivariate Statistics for the Social Science*. 5th ed. New York, NY: Taylor & Francis Group; 2009.
[72] Babilon S, Khanh TQ. A new metric for memory colour preference evaluation in lighting applications – experiments, mathematical definition, and comparison with other colour rendering indices. Paper presented at: Proceedings of the 29th Quadrennial Session of the CIE; 2019:251-260; Vienna, Austria, Commission Internationale de l'Éclairage (CIE).
[73] Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
[74] Wei M, Chen S. Effects of adapting luminance and CCT on appearance of white and degree of chromatic adaptation. *Opt Express*. 2019;27(6):9276-9286.
[75] Zhai Q, Luo MR. Study of chromatic adaptation via neutral white matches on different viewing media. *Opt Express*. 2018; 26(6):7724-7739.
[76] Babilon S. On the color rendition of white light sources in relation to memory preference (Ph.D. thesis). Technische Universität Darmstadt; 2018.

AUTHOR BIOGRAPHIES

Sebastian Babilon received a PhD degree in lighting engineering from the Technische Universität Darmstadt, Germany in 2018. Since 2014 he has been working at the Laboratory of Lighting Technology in Darmstadt—now as a Postdoc—focusing on memory related color perception, human centric lighting, and the spectral optimization of multi-channel LED luminaires for improved observer preference.

Tran Quoc Khanh is University Professor and Head of the Laboratory of Lighting Technology at the Technische Universität Darmstadt, Germany. He graduated in Optical Technologies and obtained a PhD degree in lighting engineering from the Technische Universität Ilmenau, Germany. Before being appointed as a professor, he gathered industrial experience as a project manager at Arnold and Richter Cine Technik AG in Munich, Germany. His research interests cover all important aspects of modern lighting technology including LEDs, colorimetry, human centric lighting, mesopic vision, glare, photometry, and color science-related topics.

How to cite this article: Babilon S, Khanh TQ. Impact of the adapted white point and the cultural background on memory color assessments. *Color Res Appl*. 2020;45:803–824. https://doi.org/10.1002/col.22525