Ergonomics, 2013
http://dx.doi.org/10.1080/00140139.2013.819940

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

Comparing colour discrimination and proofreading performance under compact fluorescent and halogen lamp lighting

Susanne Mayr*, Maja Köpper and Axel Buchner

Institut für Experimentelle Psychologie, Heinrich-Heine-Universität, Düsseldorf, Germany

(*Received 21 October 2012; final version received 17 June 2013*)

Legislation in many countries has banned inefficient household lighting. Consequently, classic incandescent lamps have to be replaced by more efficient alternatives such as halogen and compact fluorescent lamps (CFL). Alternatives differ in their spectral power distributions, implying colour-rendering differences. Participants performed a colour discrimination task – the Farnsworth–Munsell 100 Hue Test – and a proofreading task under CFL or halogen lighting of comparable correlated colour temperatures at low (70 lx) or high (800 lx) illuminance. Illuminance positively affected colour discrimination and proofreading performance, whereas the light source was only relevant for colour discrimination. Discrimination was impaired with CFL lighting. There were no differences between light sources in terms of self-reported physical discomfort and mood state, but the majority of the participants correctly judged halogen lighting to be more appropriate for discriminating colours. The findings hint at the colour-rendering deficiencies associated with energy-efficient CFLs.

Practitioner Summary: In order to compare performance under energy-efficient alternatives of classic incandescent lighting, colour discrimination and proofreading performance was compared under CFL and halogen lighting. Colour discrimination was impaired under CFLs, which hints at the practical drawbacks associated with the reduced colour-rendering properties of energy-efficient CFLs.

Keywords: compact fluorescent lamp; halogen lamp; colour discrimination; proofreading; colour-rendering index

1. Introduction

The days of classic incandescent lamps as the primary light sources in our households are numbered. For example, under Directive 2005/32/EC – a framework for the setting of ecodesign requirements for energy-using products – the commission of the European Union has adopted Regulation No 244/2009 which basically implies the stepwise withdrawal of incandescent household lamps from circulation. In detail, the regulation makes demands on the efficacy of a lamp, with efficacy being the quotient of the luminous flux emitted by the lamp divided by the power consumed. Lamps that do not fulfil these requirements must not be marketed. The transition process has started in September 2009 and will reach its final state in September 2016 with the ban on essentially all household lamps that perform below efficiency class B. Similar regulations aiming to reduce energy consumption have been implemented in several other countries, such as Australia (see Decision Regulatory Impact Statement, 2009) and the USA (Energy Independence and Security Act, 2007).

However, in spite of the phase-out of inefficient incandescent light bulbs, consumers can choose from a variety of alternative lamps. For example, they can choose improved incandescent bulbs containing halogen gas such as iodine and bromine (in the following named 'halogen lamp'). Although these lamps vary in technical details and, as a consequence, in energy efficiency, they are all more efficient than the conventional incandescent lamp.[1] As classic incandescent lamps, halogen lamps irradiate in the visible as well as the infrared wavelength range due to the thermal radiation of a tungsten filament. Compact fluorescent lamps (CFLs) – an even more efficient alternative – are low-pressure gas discharge lamps consisting of a glass tube with a phosphor coating of its inner surface, a special gas filling as a discharge carrier, and two electrodes. When the gas discharges, ultraviolet radiation is generated. The phosphor coating absorbs the ultraviolet radiation and converts it into radiation in the visible spectrum (Hofmann and Rasch, 2001). Finally, the most energy saving option are light emitting diodes (LED). However, so far this lamp technology does not play an appreciable role in household lighting and, hence, was not in the focus of interest of the present study.

Central to the present study are the differences between the two most prevalent replacements of classic incandescent lighting and their possible consequences on human perception and performance. The most obvious difference between improved incandescent lighting – represented by halogen lamps – and CFLs concerns the lamps' spectral power distributions. Whereas incandescent light sources can be characterised by a continuous distribution across the whole visible spectrum and beyond, CFLs have a discontinuous distribution with several spectrally narrow peaks that equal the spectral

*Corresponding author. Email: susanne.mayr@hhu.de

© 2013 Taylor & Francis

emission characteristics of the different phosphors in the lamp (see Figure 1, where the spectral distributions of the lamps used in the present study are depicted). Note that different CFLs may vary in the relative proportions of phosphors and activator additives, and they may also contain different types of phosphors. This implies that not all CFLs have identical spectral power distributions (Hofmann and Rasch, 2001; van Broekhoeven, 2001).

Closely related to the spectral power distribution are two properties that are used to specify light sources: the colour temperature and the colour-rendering index (CRI). Light sources differ in whether they tend to have a 'warm', yellowish appearance or a 'cold', white or even bluish light colour. Colour temperature is used to describe the colour of light sources. The temperature (in kelvin, K) of a black body which radiates in a certain colour is used to characterise the light colour of a lamp that has an equivalent colour appearance. The spectral radiant power distributions of incandescent lamps with a tungsten filament are approximations of black body radiator distributions (Wyszecki and Stiles, 1982) and, therefore, incandescent lamps can be very well characterised by the concept of colour temperature. The correlated colour temperature is an extension of the concept of colour temperature to other types of light sources (such as fluorescent lamps and LEDs). Simply put, the correlated colour temperature of a specific light source is the temperature of the black body that radiates in the colour closest to the colour of the light source (Wyszecki and Stiles, 1982). Incandescent light is characterised by a temperature below 3000 K which is associated with a warm colour appearance. Light of a temperature below 3300 K is called 'warm white' (according to European Standard, 2011). CFLs come with very different correlated colour temperatures as the light colour can be adjusted by the relative proportion of the different phosphors inside the glass tube.

A measure developed to characterise the colour-rendering property of a light source is the CRI (see CIE 1995). The general CRI (R_a) reflects the deviation in colour appearance of eight standardised colour samples when illuminated with the light source of interest compared to a reference illuminant. The reference illuminant is either a black body for light sources below 5000 K or, for light sources above 5000 K, a spectral power distribution of daylight, with the additional restriction that the reference illuminant should have a correlated colour temperature maximally close to the test light source. By definition, the reference illuminant has a CRI of 100. This implies that incandescent lamps, being identical or close to black bodies, have CRIs of 100 or close to this. In contrast, present-day compact fluorescent household lamps usually achieve colour-rendering indices between 80 and 89.[2]

The CRI is the only information concerning colour rendering of a light source available to the consumer. However, the appropriateness of this index has been called into question (Boyce, 1976; Guo and Houser, 2004; Rea and Freyssinier-Nova, 2008). The choice of incandescent light and daylight as the reference illuminant is based on familiarity – we are used to the illumination of objects by these illuminants and, hence, perceive the colour of objects under these light sources as natural. However, this choice implies that the colour rendering of incandescent lighting has been made the golden standard against which all other lamps are compared. While it might be adequate to compare a lamp against the most familiar lighting conditions when this lamp is to be described in terms of how natural objects appear under its lighting, there are other aspects of colour rendering which should also be captured by an appropriate index, such as the ability of the light source to represent small differences in hue between objects to enable colour discrimination (Guo and Houser, 2004; Rea and Freyssinier-Nova, 2008).

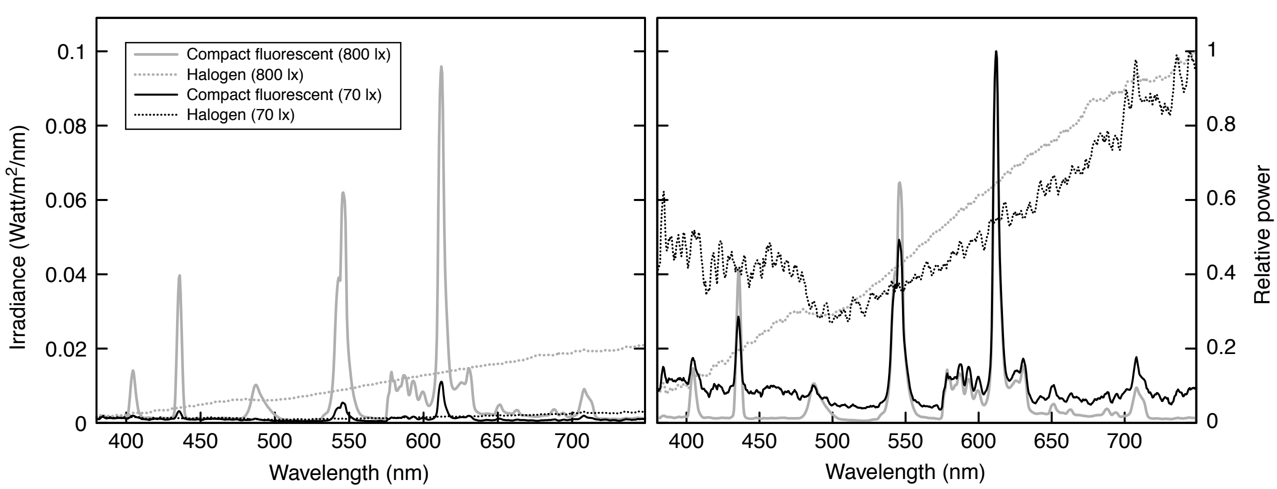

Figure 1. Irradiance (left) and relative spectral power distributions (right) of the lamps used in the present experiment as a function of light source and illuminance level. Note that the low-wattage lamps (particularly the 70 lx halogen lamp) are associated with very jagged-looking relative power distributions that are the result of inflated measurement errors.

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

Empirical evidence shows that the CRI is of only limited suitability in characterising this property of a light source. Boyce (1976) compared colour discrimination under four fluorescent lamps of strongly varying colour-rendering indices (with CRIs between 95 and 9). To this end, participants performed the Farnsworth–Munsell 100 Hue Test (Farnsworth, 1957), a test originally designed to assess human colour discrimination ability but also used to compare colour-rendering properties of different light sources. In this test, participants have to arrange colour caps of small colour differences to form a gradual transition in hue. The magnitude of transposition errors of the caps is an inverse function of colour discrimination performance. By and large, a higher CRI in the study by Boyce (1976) was associated with better performance, but depending on hue sector CRI could not reliably predict discrimination performance.

More recently, Rea and Freyssinier-Nova (2008) measured colour discrimination performance using the Farnsworth–Munsell Hue 100 Test under four LED and fluorescent lamps of CRIs between 71 and 95. Their data – though not statistically analysed – descriptively showed that CRI is not a satisfying index to describe the colour discrimination properties because there was no consistent positive relation between the CRI and colour discrimination performance. Due to the shortcomings of the CRI as a predictor of colour discrimination performance, several researchers have called for other metrics or a combination of different metrics as a better solution to cover the different aspects of colour rendering (Guo and Houser, 2004; Rea and Freyssinier-Nova, 2008).

Given the limited suitability of the CRI as an index of colour discrimination performance, it is impossible to infer whether a difference in CRI between 80 and 89 (typical for present-day standard household CFLs) on the one hand and of 100 or slightly below (achieved by incandescent lamps) on the other hand implies any measurable differences in colour discrimination performance. We thus need a direct empirical test to decide whether the two rivals in present-day household lighting differ with respect to their effects on colour discrimination performance. In the experiment reported here, we thus directly compared colour discrimination performance under halogen and CFL lighting using the Farnsworth–Munsell 100 Hue Test. Similar to the study by Rea and Freyssinier-Nova (2008, Experiment 1 and 2), we focussed on light sources of similar correlated colour temperature, excluding this variable as an explanation of possible performance differences.

Rea and Freyssinier-Nova (2008) not only tested the colour discrimination performance under different lamps, but also manipulated the illuminance level of the light sources. Participants performed the Farnsworth–Munsell 100 Hue Test either under 54 lx or 540 lx illumination. A high illuminance level facilitated colour discrimination performance under all light sources. Similarly, Knoblauch et al. (1987) found evidence of impaired colour discrimination performance at lower illuminances between 5.7 lx and 1800 lx. However, a beneficial effect of illuminance level on colour discrimination has not always been found. For example, in the study by Boyce (1976) performance in the Farnsworth–Munsell 100 Hue Test was not affected by illuminance level (300 lx vs. 1000 lx). Note that the dim level in the study by Boyce was much higher than in the two other studies which found an effect of illuminance level. It is thus possible that the illuminance level needs to be rather low to result in a measurable drop in colour discrimination performance.

In the present experiment, we manipulated illuminance level in addition to the type of light source. We were interested not only in whether illuminance level would affect colour discrimination performance *per se* – which would be a replication of the work by Rea and Freyssinier-Nova (2008) and Knoblauch et al. (1987) – but also in whether the level of illuminance would modulate the potential effects of light source. If, for example, the use of CFLs compromised colour discrimination only under dim levels of illuminance, the practical solution would be to avoid CFLs only in contexts with low levels of illuminance. We manipulated the illuminance level in two steps (70 lx vs. 800 lx). The choice of levels was motivated by the European lighting standard for indoor work places (European Standard, 2011) that recommends illuminance levels of about 500 lx for office work places.[3] The brighter illuminance level of 800 lx was chosen to be clearly above this requirement, the lower illuminance level of 70 lx was clearly below it.

The small changes in colour appearance of the CFL lighting compared to the halogen lighting (as evidenced by the slight differences in correlated colour temperature and in chromaticity coordinates; see Table 1 for the exact measurements of our lamps) might also have unspecific effects on information processing and, hence, on test performance via mood or motivation (e.g., if one light source appeared more pleasing or more familiar than the other). In order to assess whether the expected differences in the colour discrimination test were caused by the colour-rendering differences only or whether they were the result of unspecific effects of lighting differences, we also ran a colour-insensitive achromatic proofreading task. Effects of the light source manipulation in the colour discrimination task but not in the achromatic proofreading task would indicate discriminant validity of the manipulation. To further evaluate the quality of the effects found, we also assessed subjective well-being and asked participants whether they had perceived differences in lighting between the two light source conditions.

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

Table 1. Description of light sources including associated illuminance level, correlated colour temperature, chromaticity coordinates, and colour-rendering index (R_a).

Light source	Illuminance on the work plane (in lx)[a]	Correlated colour temperature (K)	Chromaticity coordinates 1931 CIE 2° (x,y)	Chromaticity coordinates 1976 CIE 2° (u',v')	Colour-rendering index (R_a)	TES[b]
Compact fluorescent (15 W)	805	2693	0.453, 0.397	0.264, 0.521	82	108 (49)
Compact fluorescent (5 W)	70	2423	0.479, 0.407	0.277, 0.529	83	151 (47)
Halogen (70 W)	801	2743	0.447, 0.392	0.262, 0.518	98	84 (46)
Halogen (18 W)	66	2737	0.450, 0.398	0.262, 0.521	98	130 (53)

Note: For measurement details, see text. Calculations are based on macros developed by the National Institute of Standards and Technology, Version 7.4 (Davis and Ohno, 2005; Ohno, 2004).[a] Mean illuminance levels – measurements were made after each change in lamp-to-chamber assignment – differed only marginally from the predefined values of 70 and 800 lx. For the sake of convenience, we will refer to the predefined values in the following.
[b] TES, mean total error scores and standard deviations (in parentheses) of the Farnsworth–Munsell 100 Hue Test.

2. Method

2.1. Design

The experiment comprised a two-factorial design with light source (CFL vs. halogen) and illuminance level (70 lx vs. 800 lx) as between-subjects variables. Every participant was consecutively tested under CFL and halogen lighting but the second testing was done for the sole purpose of yielding preference ratings based on the direct comparison between light sources. Performance measures from the second testing were not analysed because, due to carry-over effects from the first testing, they would necessarily be ambiguous and impossible to interpret. As a consequence, the light source variable was treated as a between-subjects variable with respect to the first light source tested.

2.2. Participants

Participants were 105 adults (23 males) with 25 participants in the 70 lx/CFL condition, 24 participants in the 70 lx/halogen condition, 28 participants in the 800 lx/CFL condition, and 28 participants in the 800 lx/halogen condition. Participants ranged in age from 18 to 40 years ($M = 24.49$). All except four were students. Participants received partial course credit or monetary compensation for their participation. All participants were native German speakers and reported normal or corrected-to-normal visual acuity and normal colour vision. Two participants changed their position during the colour discrimination task and abstained from using the chin rest. Colour discrimination data for one participant were not registered correctly. Another participant misunderstood the proofreading instructions and did not mark errors. One participant did not answer the questionnaire to assess physical discomfort. The missing or irregular data points were treated as missing values only for the respective dependent measures which implies varying degrees of freedom for the different analyses reported.

2.3. Apparatus and lighting conditions

Testing was conducted in two adjacent identically built and furnished testing chambers (of 120 × 225 × 250 cm width × length × height) without any exposure to daylight or any external lights other than the manipulated light sources. Both chambers were furnished with a 100 × 100 cm table with a light grey surface and a height-adjustable office chair. One of the chambers was illuminated by the CFL, the neighbouring chamber was illuminated by the halogen lamp of the same illuminance level. The illuminance level (70 lx vs. 800 lx) was changed every second or third day in order to achieve approximately equal sample sizes for both levels at the end of the data acquisition period. Moreover, the assignment of light sources (CFL vs. halogen) to the two chambers was counterbalanced (see Figure 2 for a depiction of the experimental set-up).

Four different lamps were tested: two CFLs of 5 W (Osram Dulux Superstar Mini Globe, OSRAM AG, München, Germany) and 15 W (Osram Duluxstar Mini Ball) input power, respectively, and two halogen lamps of 18 W (Osram Halogen Classic A FR ES) and 70 W (Osram Halogen Eco Classic A) input power, respectively (see Figure 1 for the spectral power distributions of these lamps; see Table 1 for the description of the light sources in terms of illuminance level on the work plane, correlated colour temperature, chromaticity coordinates, and CRI [R_a]). Measurements were made with a StellarNet BLUE-Wave Spectrometer (StellarNet Inc., Tampa, FL, USA) which has an optical resolution of 2.5 nm and an accuracy of less than 0.25 nm (settings: 135 ms detector integration time, averaged across five scans, measurement increment of 0.5 nm, temperature compensation on). The CRI of the two CFLs was in the lower 80s range (82 and 83, respectively), the CRI of 98 for the two halogen lamps was close to the maximum of 100. All four lamps were within the

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

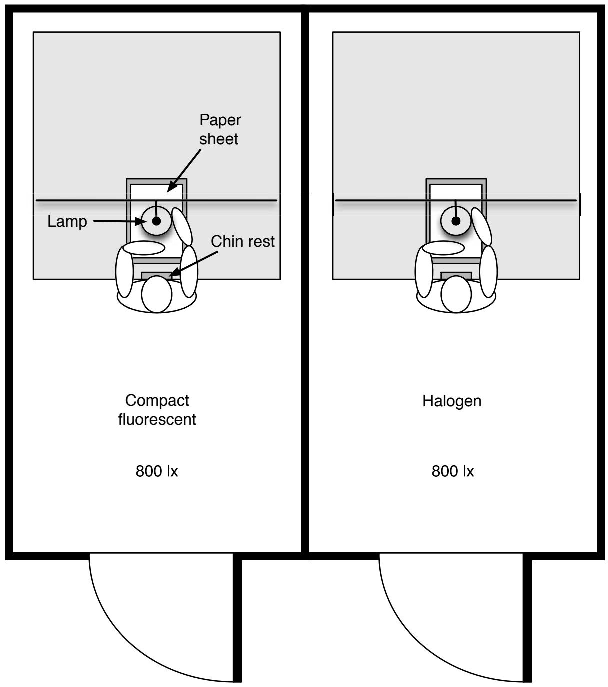

Figure 2. Depicted is the experimental set-up. At a specific day, illuminance level was always equal in both testing chambers (either 800 lx as in the depicted example or 70 lx), while light sources always differed between chambers (compact fluorescent or halogen). Illuminance level as well as the assignment of light sources to testing chambers were varied between days. Participants were tested under both light sources by changing chambers. The sequence of testing (first compact fluorescent, then halogen, or vice versa) was randomly assigned to participants.

warm white colour temperature range. A closer match in correlated colour temperature was not possible given the available product lines. All lamps were of the bulb shape type and had a standard E27 screw base. Lamps were inserted in a matching lamp socket that was mounted on a support frame directly above the tabletop on which the experimental tasks were conducted. A domed metal lamp shade with a white inner coating was positioned above the lamp.

The illuminance level was measured using a Gossen Mavolux 5032 B luxmeter (in accordance with class B DIN 5032–7; Gossen Foto-and Lichtmesstechnik GmbH, Nürnberg, Germany). For this purpose, a 210×297 mm sheet of recycled paper (80% whiteness, the same paper that was used in the proofreading task) was positioned in portrait mode directly below the mounted lamp. Illuminance level was defined as the mean value of seven predefined measurement points on the paper sheet. The height of each lamp was adjusted to a distance that resulted in the predefined illuminance level of either 70 or 800 lx, respectively. In this way, an illuminance of approximately 70 lx was achieved by positioning the 5 W CFL and the 18 W halogen lamp about 120 cm and 94 cm, respectively, above the table surface. An illuminance of approximately 800 lx was achieved by positioning the 15 W CFL and the 70 W halogen lamp about 56 cm and 62 cm, respectively, above the table surface. All lamps were switched on 30 min before measurements and experimental testing started to warm up.

2.4. *Tasks and questionnaires*

The Farnsworth–Munsell 100 Hue Test (Farnsworth, 1957) was applied to measure colour discrimination performance. This test comprises four trays of linearly aligned colour caps. Altogether the caps cover the whole hue circle, whereas the different trays cover different hue sectors (Tray A: red to green-yellow; Tray B: green-yellow to blue-green; Tray C: blue-green to purple-blue; Tray D: purple-blue to red). The two anchor caps of each tray are fixed. The 22 removable caps in Tray A and the 21 removable caps in Trays B, C, and D have to be arranged by the participant in order to form a gradual transition in hue between the two anchor cap colours. Correctly sorted caps have ascending numbers – invisible to the participants – on their backside so that the summed difference between a correctly arranged cap and its two neighbouring caps equals two. Incorrectly sorted caps result in summed differences larger than two. The magnitude of these transposition errors are summed to result in a total error score which is an inverse function of the chromatic discrimination performance. Partial errors scores for the separate trays indicate the colour discrimination performance for specific hue ranges. For scoring, the

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

associated test scoring software was used. Note that in order to measure colour discrimination performance with reference to normative data (such as the norms published by Kinnear and Sahraie 2002), the test has to be administered under daylight illuminant D65. In the present study, testing was accomplished under light sources of clearly different spectral power distributions. Thus, absolute error scores cannot be interpreted. Instead, our focus of interest is on the relative differences in performance under the different light sources.

The proofreading task comprised 12 individual short stories. Two subsets of six stories were generated. Each participant proofread one subset under the lighting condition he or she was assigned to (CFL or halogen lamp). In addition, each participant also proofread the other subset under the other lighting condition (halogen lamp or CFL); the latter was necessary in order to be able to assess subjective preferences (see below). Each story comprised between 864 and 878 words and was printed as a single left-aligned paragraph (without hyphenation) on a separate 210 × 297 mm sheet of paper. Texts were presented in 8 pt Helvetica font and comprised between 51 and 58 lines of text which was 13 cm wider and about 20–22.5 cm high. Each text contained 30 errors, 22 of which were spelling errors (letter omissions, letter substitutions, transpositions of adjacent letters, and letter additions) and eight were syntactic errors (such as incorrect flection and conjugation). Searching for syntactic errors forced participants to read the texts for comprehensibility rather than to skim single words for spelling irregularities. Errors were randomly distributed across the text. For each text, proofreading accuracy was calculated as the number of correctly detected errors corrected by the number of false alarms. In addition, the number of lines read within the reading interval was used as a measure of proofreading speed.

Aside from the measuring of colour discrimination and proofreading performance, two questionnaires were administered to assess participant's self-reported physical discomfort and mood state. The short form of the multidimensional mood state questionnaire (Steyer et al. 1997) was used to assess the subjective well-being of participants with respect to three bipolar dimensions: pleasant versus unpleasant mood, alertness versus sleepiness, and calmness versus restlessness. For each dimension several adjectives (such as 'content', 'tired', 'relaxed') had to be rated on a five-point scale (from 'not at all' [1] to 'very much' [5]) as describing the participant's current mood state. The questionnaire is available in two parallel versions, A and B. Each participant completed both the versions, one after working under CFL lighting and one after working under halogen lamp lighting. The sequence in which the two versions of the questionnaire were answered was counterbalanced. In the questionnaire to assess physical discomfort (Heuer et al. 1989), the participants' estimation of experiencing eye strain (seven items), headache (three items), and musculoskeletal strain (four items) was measured by describing physical states of discomfort (such as 'I have burning eyes'). The participants had to rate (from 'not at all' [1] to 'very much' [7]) the extent to which the description was representative of their own state.

In a final questionnaire, participants were asked whether they had noticed a difference in illumination between the two testing chambers (Question 1), whether they had preferred the illumination of one of the chambers over the other in general (Question 2), whether they had preferred the illumination of one of the chambers over the other for the proofreading task (Question 3) or for the colour discrimination task (Question 4). Answer options were 'Chamber 1', 'Chamber 2', and 'no difference'. Finally, there was free space for any comments to the experiment.

2.5. Procedure

Participants were tested individually. They were naive as to the purpose of the study. Participants were seated at the table in one of the two chambers. They were allowed to adjust the height of the chair in order to sit comfortably while positioning their chin on a chin rest at a height of 42 cm above the desktop. The chin rest prevented participants from bending over the workspace which would have resulted in blocking the experimentally manipulated lighting from above. This position allowed all participants regardless of body size to comfortably circle erroneous words in the proofreading task and to arrange the colour caps in the colour discrimination task. Participants adapted to the lighting condition in the first chamber for about three minutes while receiving task instructions from the experimenter. Testing started with the proofreading task. An auditory signal presented over headphones prompted participants to start and stop proofreading of the six texts. Reading intervals were 3 min for each text. Auditory cues appeared at 1:30 min and at 2:50 min, informing participants about the time remaining for a particular text. The proofreading task was followed by the colour discrimination task. The experimenter explained the task and showed a picture of the ordered trays. For testing, the four trays were always presented in the same sequence (A, B, C, and D). Each tray has a protective cover which was used to present the caps. The caps of each tray were shuffled by the experimenter and then aligned in random sequence into the cover of the respective tray. The cover containing the random sequence of caps together with the empty tray (solely including the two anchor caps) were positioned on the table in front of the participants. Participants were given a signal to start the sorting task. After two minutes, participants were reminded to finish the task soon, but all participants were allowed to complete the task. Finally, the multidimensional mood state questionnaire and the physical discomfort questionnaire were administered. Then, the participant left the first chamber and moved over to the adjacent chamber illuminated with the other light source but the same illuminance level. The same sequence of tests and questionnaires immediately started. Finally, participants left the second chamber and completed the preference questionnaire in a room lit by daylight without any extra lighting sources. After completing the preference questionnaire, all participants were informed about the purpose of the experiment. The experiment lasted about 75 min.

2.6. *Dependent measures and statistical power*

The dependent variables were participants' total and partial error scores and their average processing times in the Farnsworth–Munsell 100 Hue Test, proofreading accuracy (P_r = hits − false alarms) and proofreading speed (number of lines read), the ratings of the three scales in the mood state questionnaire and of the three scales in the physical discomfort questionnaire, and the response frequencies of the four questions in the preference questionnaire.

In order to detect large effects of size $f = 0.40$ (as defined by Cohen, 1988) for the light source and the illuminance level manipulation, given desired levels of $\alpha = \beta = 0.05$, data had to be collected from a sample of at least $N = 84$ participants (Faul et al. 2007). We were able to collect data from $N = 105$ participants so that the power was $(1 - \beta) = 0.98$ and thus even larger than what we had planned for. The level of alpha was maintained at 0.05 for all statistical decisions.

3. Results

3.1. *Colour discrimination performance*

The means of participants' colour discrimination performance in the Farnsworth–Munsell 100 Hue Test are presented in Figure 3. With respect to the total error scores (leftmost columns), participants made more transposition errors under dim illumination conditions. For both illumination conditions, the number of errors was larger with CFL lighting.

A two-factorial ANOVA of the total error score data with illuminance level (70 lx vs. 800 lx) and light source (CFL vs. halogen) as between-subjects variables revealed a significant effect of illuminance level, $F(1, 98) = 21.18$, $p < 0.01$, $\eta^2 = 0.18$, and a significant effect of light source, $F(1, 98) = 5.49$, $p = 0.02$, $\eta^2 = 0.05$. The interaction between both variables was not significant, $F(1, 98) < 1.00$.

In order to reveal whether the decline in colour discrimination performance with CFL lighting was specific for certain sections in colour space, we ran the same analysis for each of the four trays (right columns of Figure 3). For Tray A (covering the red to green-yellow sector), the two-factorial ANOVA of the total error score data revealed a significant effect of illuminance level, $F(1, 98) = 21.98$, $p < 0.01$, $\eta^2 = 0.18$, with more transposition errors under dim (70 lx) than under bright (800 lx) illumination. Neither the main effect of light source, $F(1, 98) = 1.33$, $p = 0.25$, $\eta^2 = 0.01$, nor the interaction were significant, $F(1, 98) = 1.04$, $p = 0.31$, $\eta^2 = 0.01$.

For Tray B (green-yellow to blue-green), there was also a significant effect of illuminance level, $F(1, 98) = 17.81$, $p < 0.01$, $\eta^2 = 0.15$, with worse performance under dim than under bright illumination. In addition, participants made more errors under CFL than under halogen lighting, $F(1, 98) = 38.98$, $p < 0.01$, $\eta^2 = 0.29$. The interaction between both variables was not significant, $F(1, 98) < 1.00$.

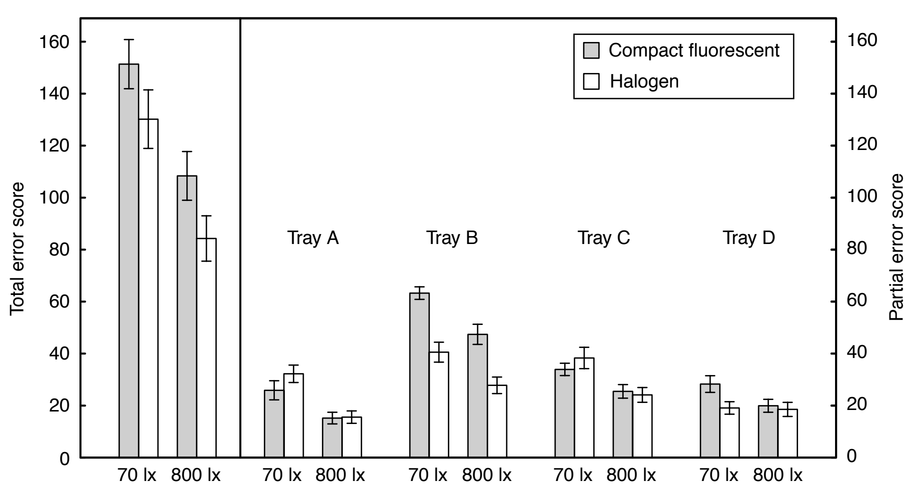

Figure 3. Total (left) and partial (right) error scores in the Farnsworth–Munsell 100 Hue Test as a function of light source and illuminance level. The error bars depict the standard errors of the means.

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

For Tray C (blue-green to purple-blue), the pattern of results was similar to Tray A. There was only a significant increase in transposition errors under dim illumination, $F(1, 98) = 14.37$, $p < 0.01$, $\eta^2 = 0.13$. Neither the effect of light source, $F(1, 98) < 1.00$, nor the interaction were significant, $F(1, 98) < 1.00$.

Finally, for Tray D (purple-blue to red), the errors increased under CFL lighting but this effect just missed the preset level of significance, $F(1, 98) = 3.71$, $p = 0.06$, $\eta^2 = 0.04$. No other effects were significant, $F(1, 98) = 2.63$, $p = 0.11$, $\eta^2 = 0.03$, for the effect of illuminance level, and $F(1, 98) = 2.02$, $p = 0.16$, $\eta^2 = 0.02$, for the interaction effect.

Analogous analyses of processing time revealed for Tray B that participants needed more time to finish the task under CFL lighting ($M = 100.33$ s, $SE = 3.68$) as compared to halogen lamp lighting ($M = 89.88$ s, $SE = 3.32$), $F(1, 98) = 4.46$, $p = 0.04$, $\eta^2 = 0.04$. This finding implies that the increase in transposition errors under CFL lighting for Tray B was not compromised by a speed-accuracy trade-off. No other effects were significant, all remaining $F(1, 98) < 1.28$, $p > 0.26$.

3.2. Proofreading performance

The means of participants' proofreading accuracy are presented in Figure 4 (upper panel). Participants performed worse under dim (70 lx) than under bright (800 lx) lighting. A two-factorial ANOVA of the accuracy data confirmed this effect of

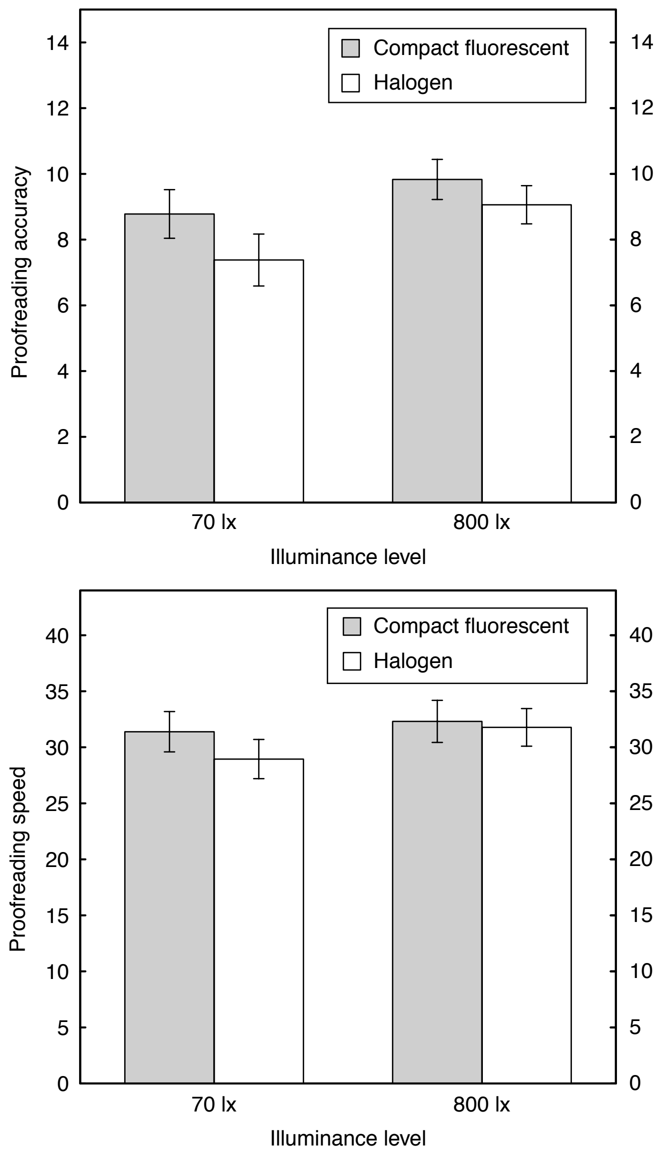

Figure 4. Performance in the proofreading task as a function of light source and illuminance level. The upper panel depicts proofreading accuracy measured as the average number of correctly detected errors corrected by the number of false alarms. The lower panel depicts proofreading speed measured as the average number of lines read. The error bars depict the standard errors of the means.

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

illuminance level, $F(1, 100) = 4.07$, $p = 0.05$, $\eta^2 = 0.04$. Neither the effect of light source, $F(1, 100) = 2.56$, $p = 0.11$, $\eta^2 = 0.03$, nor the interaction were significant, $F(1, 100) < 1.00$. The analogous analysis of proofreading speed (Figure 4, lower panel) revealed no significant effects, all $F(1,100) < 1.11$, $p > 0.29$.

3.3. *Mood state and physical discomfort ratings*

Mood state measurements did not vary as a function of any of the independent variables, all $F(1, 101) < 3.73$, $p > .05$ (Figure 5, upper panel). Similarly, the physical discomfort ratings (Figure 5, lower panel) did not generally vary as a function of any of the independent variables, all $F(1, 100) < 3.50$, $p > .06$. There was one exception in that a significant effect of illuminance level was found for the headache ratings, $F(1, 100) = 4.49$, $p = 0.04$, $\eta^2 = 0.04$. Participants reported more problems under dim than under bright lighting.

3.4. *Preferences*

Of the 105 participants, only 39 claimed to have noticed a difference in the lighting between the two testing chambers (Question 1). Obviously, the lighting difference between chambers had been fairly subtle. For each of the following preference questions three answers were possible ('Chamber 1', 'Chamber 2', and 'no difference'), but 'no difference'

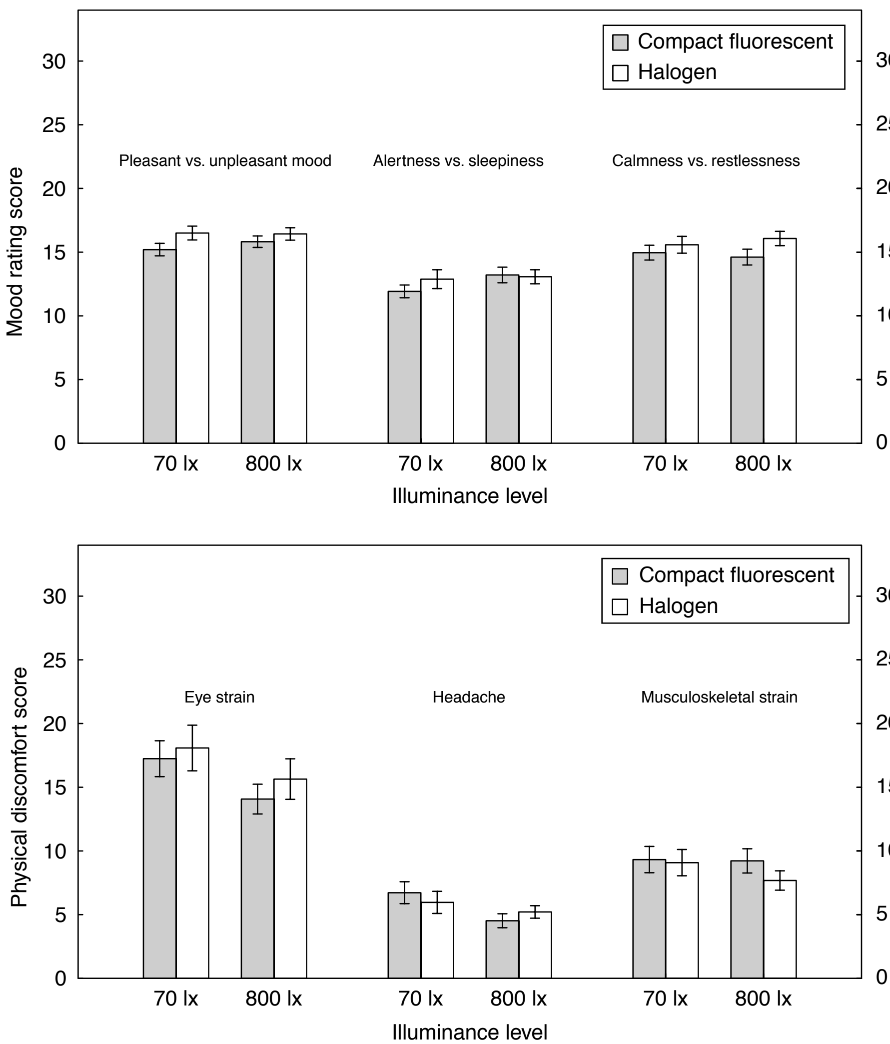

Figure 5. Subjective well-being measures as a function of light source and illuminance level. The upper panel depicts the ratings of the multidimensional mood state questionnaire. Larger scores indicate a more positive mood rating. The lower panel depicts the ratings of the questionnaire to assess physical discomfort. Larger scores indicate a stronger degree of physical discomfort. The error bars depict the standard errors of the means.

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

responses were excluded from the analyses. There was no general preference for one of the two chambers with respect to lighting quality (Question 2: 45 vs. 44 for CFL vs. halogen), $\chi^2(1) = 0.01$, $p = 0.92$. The same was true when participants were asked for their lighting preference for the proofreading task (Question 3: 40 vs. 48 for CFL vs. halogen), $\chi^2(1) = 0.73$, $p = 0.39$. However, for the colour discrimination task, there was a significant preference for the chamber with halogen lamp light (Question 4: 31 vs. 50 for CFL vs. halogen), $\chi^2(1) = 4.46$, $p = 0.04$. In summary, participants had not detected a general difference in lighting between the chambers but they were sensitive enough to notice that the colour discrimination task was easier to accomplish under halogen than under CFL lighting.

4. Discussion

Arguments about the energy efficiency of CFL as compared to incandescent lighting have led to legislation that will essentially ban classic incandescent lighting from our households. A remaining choice is that between CFL and improved incandescent (i.e., halogen) lamps. Although the differences in spectral power distribution of these two lamp types are well known, it is still unclear whether they are of any relevance for human perception and performance. To this end, we compared colour discrimination performance under halogen and CFL lighting and, additionally, varied illuminance level as a potentially moderating factor. In order to assess whether the expected differences in the colour discrimination test were caused by the colour-rendering differences only or whether they were the result of unspecific effects of lighting differences, we also ran a colour-insensitive achromatic proofreading task.

The pattern of results is very clear. Compared to incandescent halogen lighting, CFL lighting impairs colour discrimination performance. This was true regardless of illuminance level. The loss in colour discrimination performance under CFL lighting was so strong that the majority of participants were subsequently able to tell which of the two testing chambers – that differed only with respect to the source of lighting – was more suitable for accomplishing the colour discrimination task. Note that this retrospective sensitivity in judging the lighting quality between chambers is not at all trivial because of two reasons. First, participants did not receive feedback about the correctness of how they arranged the colour caps. As a consequence, the increased difficulty in sorting the colour caps under CFL lighting must have been fairly strong. Second, participants had not been informed before testing that light sources were manipulated between the two testing chambers so that it is unlikely that they were particularly attentive to the lighting differences when sorting the caps. In line with this second aspect, the first question of the final questionnaire revealed that only a minority of participants noticed a general difference between lighting in the two chambers. The combined response pattern indicates that the overall lighting difference between the chambers was so subtle that most participants did not notice it. Notwithstanding this insensitivity for the differences in illumination, the difference in colour discrimination difficulty between the chambers was noticed by the majority of participants.

Similar to the present study, Royer, Houser, and Wilkerson (2012) compared performance in the Farnsworth–Munsell 100 Hue Test under four different light sources. Incandescent halogen lamplight and two types of fluorescent lamps were compared with a RGB LED of similar correlated colour temperature, which was also close to the warm white light colour used in the present study. Under LED lighting, participants committed more transposition errors than in any other condition, while there was no difference in colour discrimination performance between halogen and CFL lighting. The latter finding differs from that of the present study. The reason for this discrepancy is not obvious. It is possible that the slight differences in spectral power distribution, correlated colour temperature, and CRI between the fluorescent lamps in the two studies are the reason for the differences in results (for details, please compare Table 1 and Figure 1 in Royer, Houser, and Wilkerson 2012, with Table 1 and Figure 1 of the present study).

Another possibility may be related to the fact that Royer, Houser, and Wilkerson (2012) used a within-subject manipulation of light sources. Within-subject designs are prone to expectancy biases, with participants developing opinions throughout the course of the experiment about the comparative difficulty and/or convenience of the different conditions as well as about the underlying experimental hypotheses. Veitch and McColl (2001; see also Veitch, Gifford, and Hine 1991 for a direct manipulation of expectancies) have argued that expectancy effects might be of particular relevance in lighting research where within-subject designs are prevalent and the manipulation (of light source) is usually visible to the participant. Buchner and Baumgartner (2007) have argued that within-subject designs are prone to a reduced probability of finding effects between experimental conditions of different task difficulty because participants may wish to perform at a certain standard and thus invest more effort in difficult than in easy tasks.[4] This reduces differences between conditions such that, in the Royer, Houser, and Wilkerson's study, only the largest difference between LED and all other lighting conditions may have become visible. Of course, these considerations are entirely post hoc. Future studies – that should also include LEDs as the currently most promising light source technology – will be necessary to resolve this issue.

However, the data from Royer, Houser, and Wilkerson's study and the present data are consistent in that the differences in transposition errors as a function of lighting condition were most frequent in the green-yellow to blue-green range covered in Tray B of the Farnsworth–Munsell 100 Hue Test. This may be at least partly due to the fact that Tray B covers

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013

the most difficult part of the colour discrimination task[5] (if a task is so easy that it can be solved under any condition, then it cannot be sensitive to between-condition differences), but it may also point to the fact that colour perception in this section of the spectrum may be particularly sensitive to inhomogeneities in the emission spectra of the light sources. For example, Figure 1 shows that the spectral power distributions of the CFLs emitted almost no radiation in the range between about 505 and 530 nm and also very little between 445 and 480 nm. If there is no light emission in a certain wavelength range, no light in this range can be reflected from the colour caps and, as a consequence, the caps cannot be distinguished one from another on the basis of reflectance differences in this wavelength range.

The reduction in colour discrimination performance under CFLs was independent of the illuminance level (70 lx vs. 800 lx). This implies that lower levels of illumination do not exacerbate the colour discrimination problems under CFLs. What we have not shown is whether higher levels of illumination may reduce these problems. Note that the colour discrimination impairment was absolutely constant across a fairly large section of the illumination range that is typical for artificial lighting situations. Given this, it seems that an illumination level *much* higher than the 800 lx used here would be required before any sizeable reduction in the CFL-related colour discrimination impairment could reasonably be expected. Obviously, the higher the CFL illumination level has to be in order to allow for colour discrimination performance that is typical for much lower halogen illumination levels, the smaller the energy efficiency advantage of CFLs over halogen lamps.

While there was a clear disadvantage for CFL lighting in the colour discrimination task, none of the other measures revealed a difference between the light sources. As expected, proofreading performance was accomplished equally well under halogen and CFL lighting. Also, none of the subjective measures of well-being (mood states and physical discomfort) differed between light sources. These results suggest that the effects in colour discrimination found in the present study trace back to the colour-rendering differences of the light sources and are not the result of rather unspecific and subtle changes in colour appearance (with respect to colour temperature and chromaticity coordinates; see Table 1) between CFL and halogen lighting. Furthermore, these results are compatible with the assumption that energy-efficient CFL lighting is equivalent to incandescent lighting as long as colour perception is irrelevant.

Compared to the manipulation of light sources, the manipulation of illuminance level had much broader consequences. First, colour discrimination was significantly impaired under dim illumination compared to bright illumination overall, but also in three of the four hue ranges tested (i.e., Trays A, B, and C). This replicates the findings by Rea and Freyssinier-Nova (2008) and Knoblauch et al. (1987). Second, proofreading performance suffered when lighting was poor. Third, participants reported stronger symptoms of headache under dim levels of illuminance. Not surprisingly, bright illumination levels are advantageous for task performance and subjective well-being. Strictly speaking, our manipulation of illumination level was confounded with the spectral power distribution of the lamps because we used different lamps for the bright and the dim illumination levels. While using a mechanical filter would have been the experimentally more elegant solution to manipulate illuminance level without changing lamps (and, hence, without changing their spectral power distributions), we chose the ecologically more valid solution. If a consumer wants to buy a new light source that is equivalent to the present room illumination but somewhat brighter, he/she would likely also buy a lamp from the same manufacturer, with the same or the most similar CRI (range) and correlated colour temperature specifications, but with more luminous flux. Although the spectra of the dim and bright halogen lamps differed slightly, their properties were almost identical (identical CRIs and almost identical colour temperatures and chromaticity coordinates, see Table 1). Similarly, the two CFLs differed only slightly in their properties. In the performance data, none of our analyses revealed an interaction between light source and illumination level. Thus, we conclude that the difference in spectral power distribution between the bright and dim lamps did not have any measurable influence on the illumination level effect.

5. Conclusion

The present study revealed impaired colour discrimination performance under CFL lighting as compared with incandescent halogen lighting, regardless of illuminance level. The disadvantage of CFL lighting was limited to the colour-sensitive discrimination task; achromatic proofreading performance as well as measures of subjective well-being were not affected by the different light sources. Thus, CFL lighting may be an adequate replacement for incandescent lamps when colour perception is irrelevant. However, even small decreases in colour-rendering quality – such as a drop of CRI from almost 100 to a value in the range 80–89 – go along with noticeable decreases in colour discrimination performance. The relatively high efficiency associated with CFLs is only one side of the coin. The other side of the coin – the reduction in colour discrimination performance – is usually neglected. It is a social rather than a scientific question whether we want to see and discriminate fine gradations in colour in the world around us, or whether we are willing to sacrifice some of this colour discrimination in order to save energy.

Acknowledgements

The authors would like to thank Eva-Louisa Lamerz and Inga Lewald for their assistance with data collection.

Notes

1. While improved incandescent halogen lamps with xenon gas filling fulfil class C of the energy label and will be marketed only until 2016, improved incandescent halogen lamps with infrared coating conform with class B of the energy label and can be marketed beyond 2016. For details about the phase-out process in Europe and for numerical comparisons of energy efficiency between lamp types, see the FAQ at the official information website of the European Commission's Directorate-General for Energy (2009), http://ec.europa.eu/energy/lumen/index_en.htm.
2. Depending on the types and concentrations of phosphors and activator additives included in a lamp, CFLs achieve very different CRIs that can also be above or below the range of 80–89.
3. There are no equivalent standards for household lighting. Given that there is a considerable overlap between typical office work and work accomplished under household lighting (e.g., paperwork at the writing desk at home), a reference to an indoor work place norm seemed adequate.
4. Counterbalancing order in a within-participants design – as done by Royer, Houser, and Wilkerson (2012) by using a Latin square design – is an appropriate means to control for simple order effects (e.g., participants being generally better in the second condition due to training), but it cannot exclude interaction effects between the light source and the order variable.
5. Tray B was associated with the largest partial error score of all four trays (see Figure 3).

References

Boyce, P. R. 1976. "Illuminance, Lamp Type and Performance on a Colour Discrimination Task." *Lighting Research and Technology* 8 (4): 195–199.

Buchner, A., and N. Baumgartner. 2007. "Text-Background Polarity Affects Performance Irrespective of Ambient Illumination and Colour Contrast." *Ergonomics* 50 (7): 1036–1063. doi:10.1080/00140130701306413.

CIE (Commission Internationale de l'Éclairage). 1995. "Method of Measuring and Specifying Colour Rendering Properties of Light Sources." Technical Report No. 13.3-1995. Vienna: CIE.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

Commission Regulation (EC) No 244/2009. *OJ L 76*. p. 3–16 of 24.3.2009.

Davis, W., and Y. Ohno. 2005. "Toward an Improved Color Rendering Metric." *Proceedings of SPIE – The International Society for Optical Engineering*, 5941: 1–8. Article no. 59411G.

Decision Regulatory Impact Statement. 2009. *Proposed MEPS for Incandescent Lamps, Compact Fluorescent Lamps and Voltage Converters*. Australia: Ministerial Council for Energy.

Directive 2005/32/EC. *OJ L 191*. p. 29–58 of 22.7.2005.

Energy Independence and Security Act. 2007. Public Law 110-140. USA.

European Commission's Directorate-General for Energy. 2009. "Frequently Asked Questions on the Regulation Phasing Out Conventional Incandescent Bulbs.", http://ec.europa.eu/energy/lumen/doc/full_faq-en.pdf

European Standard. 2011. *Light and Lighting – Lighting of Work Places – Part 1: Indoor Work Places* (EN 12464-1:2011). European Committee for Standardization.

Farnsworth, D. 1957. *The Farnsworth–Munsell 100-Hue Test for the Examination of Color Discrimination*. Baltimore, MD: Munsell Color.

Faul, F., E. Erdfelder, A.-G. Lang, and A. Buchner. 2007. "G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences." *Behavior Research Methods* 39 (2): 175–191, doi: 10.3758/BF0319314617695343.

Guo, X., and K. W. Houser. 2004. "A Review of Colour Rendering Indices and Their Application to Commercial Light Sources." *Lighting Research and Technology* 36 (3): 183–199. doi: 10.1191/1365782804li112oa.

Heuer, H., G. Hollendiek, H. Kröger, and T. Römer. 1989. "Die Ruhelage der Augen und ihr Einfluß auf Beobachtungsabstand und visuelle Ermüdung bei Bildschirmarbeit [Resting position of the eyes and its influence on observation distance and visual fatigue during visual display work]." *Zeitschrift für experimentelle und angewandte Psychologie* 36 (4): 538–566.

Hofmann, D., and E. Rasch. 2001. "Fluorescent Lamps." In *Revolution in Lamps: A Chronicle of 50 Years of Progress*, edited by R. Kane and H. Sell, 161–200. Lilburn, GA: The Fairmont Press.

Kinnear, P. R., and A. Sahraie, A. 2002. "New Farnsworth–Munsell 100 Hue Test Norms of Normal Observers for Each Year of Age 5–22 and for Age Decades 30–70." *The British Journal of Ophthalmology* 86 (12): 1408–1411. doi: 10.1136/bjo.86.12.1408.

Knoblauch, K., F. Saunders, M. Kusuda, R. Hynes, M. Podgor, K. E. Higgins, and F. M. de Monasterio. 1987. "Age and Illuminance Effects in the Farnsworth–Munsell 100-Hue Test." *Applied Optics* 26 (8): 1441–1448. doi: 30235 [pii].

Ohno, Y. 2004. "Color Rendering and Luminous Efficacy of White LED Spectra." *Proceedings of SPIE – The International Society for Optical Engineering*, 5530: 88–98. Article no.12.

Rea, M. S., and J. P. Freyssinier-Nova. 2008. "Color Rendering: A Tale of Two Metrics." *Color Research and Application* 33 (3): 192–202. doi: 10.1002/col.20399.

Royer, M. P., K. W. Houser, and A. M. Wilkerson. 2012. "Color Discrimination Capability Under Highly Structured Spectra." *Color Research and Application* 37 (6): 441–449. doi: 10.1002/col.20702.

Steyer, R., P. Schwenkmezger, P. Notz, and M. Eid. 1997. *Der mehrdimensionale Befindlichkeitsfragebogen (MDBF)* [The multidimensional mood questionnaire (MDBF)]. Göttingen: Hogrefe.

van Broekhoeven, J. 2001. "Lamp Phosphors." In *Revolution in Lamps: A Chronicle of 50 Years of Progress*, edited by R. Kane and H. Sell, 93–126. Lilburn, GA: The Fairmont Press.

Veitch, J. A., R. Gifford, and D. W. Hine. 1991. "Demand Characteristics and Full Spectrum Lighting Effects on Performance and Mood." *Journal of Environmental Psychology* 11 (1): 87–95. doi: 10.1016/S0272-494428052980007-6.

Veitch, J. A., and S. L. McColl. 2001. "A Critical Examination of Perceptual and Cognitive Effects Attributed to Full-Spectrum Fluorescent Lighting." *Ergonomics* 44 (3): 255–279, doi: 10.1080/00140130010002008l1219759.

Wyszecki, G., and W. S. Stiles. 1982. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. 2nd ed. New York: John Wiley.

Downloaded by [Universitaets und Landesbibliothek] at 02:20 26 August 2013