
Published paper

Fotios, S.A. and Cheal, C. (2010) *A comparison of simultaneous and sequential brightness judgements*. Lighting Research and Technology, 42 (2). 183 - 197.
ISSN 1477-1535

http://dx.doi.org/10.1177/1477153509355506

A Comparison of Brightness Judgements Made Using Simultaneous and Sequential Modes of Evaluation

SA Fotios PhD, BEng, CEng, MEI, MSLL, MILE
C Cheal PhD, MSc, BSc
School of Architecture, University of Sheffield, UK

Fotios SA & Cheal C, A Comparison of Simultaneous and Sequential Brightness Judgements. Lighting Research & Technology, 2010; 42(2); 183-197

Abstract

This article presents the results of brightness matching and brightness discrimination tests carried out using sequential evaluation (temporal juxtaposition) to compare brightness under lamps of different spectral power distribution at mesopic levels of illumination. These data are compared with the results of previous tests which used simultaneous evaluations (spatial juxtaposition) to enable comparison of these different modes of evaluation. It is concluded that sequential and simultaneous evaluations yield similar estimates of illuminances required for equal spatial brightness and similar levels of precision in this task.

1 INTRODUCTION

The authors have previously reported the results of brightness matching and brightness discrimination tasks using simultaneous evaluations – side-by-side booths lit simultaneously by two different light sources [Fotios & Cheal, 2007]. This article presents the results of a new series of tests in which the matching and discrimination tasks were repeated using the same set of lamps as before but now using sequential evaluations – a single booth lit in temporal alternation by two different light sources.

Sequential evaluations have been used in previous work to discriminate between brightnesses under lighting of different spectral power distribution (SPD) [Berman et al, 1990; Vrabel et al, 1998]. However, Yeshurun et al suggest that the sequential forced choice discrimination task is not bias free and therefore should be used with caution, if at all [Yeshurun, Carrasco & Maloney, 2008]. Uchikawa and Ikeda [Uchikawa & Ikeda, 1986] concluded that simultaneous evaluations of brightness tend to result in more stable results than sequential evaluations and this has lead at least one research group to adopt simultaneous evaluation rather than sequential for their brightness judgements [Bullough, Yuan & Rea, 2007]. The aim of this article is to review evidence for effects of evaluation mode on judgements of spatial brightness, thus to assist analysis of past research and to guide best practice for future work.

There are two modes for evaluation of multiple stimuli, joint evaluation and separate evaluation [Hsee et al, 1999] and the joint evaluation can be further sub-divided into temporal and spatial juxtapositions. The separate evaluation means that stimuli are presented individually and test participants provide a judgement of absolute magnitude. The joint evaluation permits a stimulus to be judged in comparison with a simultaneous reference stimulus, a relative judgement; the two stimuli are presented simultaneously in spatially adjacent fields, or sequentially at the same spatial location. Joint evaluations typically seek responses by the methods of forced choice discrimination (e.g. *which stimulus is brighter?*) or adjustment (e.g. *match stimuli for equal brightness by adjusting the illuminance of one stimulus*). The spatial juxtaposition is frequently a left-right comparison but top-bottom and centre-surround comparisons have also been used. All three modes of evaluation have been used in previous research of spatial brightness. For example, at photopic levels of adaptation Fotios & Levermore presented two stimuli simultaneously in side-by-side booths and used a matching task [Fotios & Levermore, 1997], Berman *et al* used a discriminating task and presented two stimuli in rapid sequential alternation at the same location [Berman *et al*, 1990], and Boyce & Cuttle presented stimuli separately in rooms and used a category rating task [Boyce & Cuttle, 1990].

Further experimental work was carried out to enable quantitative comparison of the two joint modes of evaluation in the context of lamp spectrum effects on spatial brightness. There are two reasons which suggest simultaneous and sequential evaluations could lead to different estimates of spatial brightness under different lamps. Firstly, the two evaluation modes may differently affect operation of the part of the visual system which mediates spatial brightness above and beyond the achromatic luminance response. In this case, the illuminance ratio of two stimuli at equal brightness would be different for the two evaluation modes. Secondly, the two evaluation modes may differently affect an observer's ability to carry out the brightness matching task and will impose different types of bias. There is, for example, suggestion that simultaneous evaluations may enable more precise judgements to be made than with sequential evaluation [Uchikawa & Ikeda, 1986; Jäkel & Wichmann, 2006], and this would be seen as a smaller standard deviation.

The joint evaluation in which the two stimuli are shown simultaneously at separate spatial locations is hereafter called the ***Simultaneous*** mode of evaluation; previous studies have also labelled this as a *two-alternative* or *spatial* evaluation. The joint evaluation in which two stimuli are presented in succession, often, but not necessarily, at the same spatial location, is hereafter known as the ***Sequential*** mode of evaluation; previous studies have also labelled this as the *two-interval* or *temporal* evaluation. The sequential evaluation in which the stimuli are each presented only once is called a *successive* evaluation.

2 EXPERIMENTAL BIAS

Simultaneous and sequential evaluations may suffer from bias associated with the stimulus juxtaposition, a positional bias in simultaneous evaluations and an interval bias in sequential evaluations. The response tasks themselves invoke further sources of bias, such as biases associated with the adjustment procedure in the matching task [Fotios 2001 dimming; Fotios & Cheal 2007 r.c.b. ; Fotios, Houser & Cheal, 2008] and bias associated with stimulus range selection in the discrimination task [Fotios & Cheal, 2008; Teller, Pereverzeva & Civan, 2003].

2.1 Positional Bias

Positional bias is found with spatial juxtaposition and means that one spatial location appears brighter than the other with a greater frequency than is expected, leading to an incorrect estimate of the relative brightnesses of the stimuli. Such bias can be easily seen in null condition trials, where stimuli of the same SPD and illuminance should appear equally bright, giving an illuminance ratio (left-hand/right-hand) of 1.0 in brightness matching. However, the results from some studies reveal an illuminance ratio (LH/RH) significantly different to unity. This bias may result from the inability for the experimenter to completely match the physical arrangement of two spaces, in particular when these are large fields simulating an office environment, or it may be that the test subject considers one spatial location to be brighter than the other regardless of the stimulus as was reported by Kinney [Kinney, 1955]. Whilst positional bias has caused an illuminance difference of up to 15% in previous work [Fotios, Houser & Cheal, 2008] some studies have shown that positional bias can be negligible

[Boyce, 1977]. In anticipation of a positional bias the spatial location of stimuli in the simultaneous evaluation should be counterbalanced.

2.2 Interval Bias

In sequential evaluations stimuli are presented in temporally sequential intervals and these are usually at the same spatial location. Interval Bias [Yeshurun, Carrasco & Maloney, 2008] is a consistent asymmetry in the direction of a certain response, for example a '*brighter'* response which appears with a greater frequency than is expected, and is analogous to the positional bias of simultaneous evaluations. In their detection task, Jäkel and Wichmann found a strong bias to the second interval with successive evaluations whilst the simultaneous evaluation was virtually unbiased [Jäkel & Wichmann, 2006]. Other studies investigating a range of judgements indicate a bias toward the first interval [Yeshurun, Carrasco & Maloney, 2008].

In sequential evaluations, observers have to retain their sensory impression of the preceding stimulus in mind while waiting for and then judging the current stimulus [Jäkel & Wichmann, 2006]. Thus a possible explanation of interval bias is memory limitations: the observer either cannot or does not record an accurate sensory intensity in the first stimulus when making comparison with the second stimulus [Yeshurun, Carrasco & Maloney, 2008]. Mental representations of previously encountered physical stimuli tend to be lower (e.g. shorter in length, or less bright) than were the original stimuli [LaBoeuf & Shafir, 2006] as was found in the Uchikawa and Ikeda brightness matching results where stimuli were recalled as being darker with successive evaluation than with simultaneous evaluation [Uchikawa & Ikeda, 1986].

Zheleznikova & Myasoedova [Zheleznikova & Myasoedova, 1995] used a successive matching task to compare brightness under two different types of lamp. Test participants entered the first of two rooms which was lit by one type of lamp, and then entered the second room which was lit by a second type of lamp. The amount of light in this second room was adjusted to match the brightness of the first room. The mean illuminance ratio (room-1/room-2) was greater than unity and a first interpretation is that the second room tended to appear brighter than the first room. However, the result may also be interpreted to suggest a bias toward the second interval, or, due to lack of counterbalancing in the experimental procedure, a conservative adjustment bias.

Previous research exploring interval bias tends to have used successive evaluations, in which the stimuli are presented only once each, with judgements being made after observation of the second and without the opportunity to see the first stimulus again. Two studies of spatial brightness have used sequential evaluation rather than successive. Berman et al presented three alternations of the two lamps, each being presented for five seconds at a time with a dim period of 25ms in the 100ms changeover duration [Berman et al, 1990]. Vrabel et al presented their stimuli for three seconds each, with a two second dark interval, and observers were able to ask for as many repeat presentations as needed [Vrabel et al, 1998]. The repeated presentation of both stimuli may overcome interval bias due to memory affects as the internal brightness reference is repeatedly refreshed by observation of the external reference. However there are no null condition data within these studies with which to analyse bias effects.

2.3 Precision

Precision refers to the repeatability, or stability, of judgements, the degree to which test participants are able to make the same judgement on repeat observations of the same stimuli. Lower precision would lead to larger variance in the results.

Uchikawa and Ikeda [Uchikawa & Ikeda, 1986] compared successive and simultaneous brightness matching tasks using a two-degree bipartite field. In the simultaneous task, test and comparison stimuli were presented simultaneously for one second with two-second intervals. For the successive task the test stimulus was shown first, for one second, and then the comparison stimulus was shown repeatedly for one second with two-second intervals; this matching task hence relies on memory when adjusting the second stimulus as there is no opportunity to see the first stimulus again. Results obtained using simultaneous evaluation were more stable than those using successive evaluation, the standard deviation for the successive being 20% and 130% greater than that for the simultaneous for their two observers. Jäkel & Wichmann also found lower precision with their successive detection task than with their simultaneous detection tasks [Jäkel & Wichmann, 2006].

In contrast to this, Foster et al [Foster, Amano & Nascimento, 2001] found a lower variance between observers with sequential evaluations than with simultaneous evaluations in a colour matching task. In their sequential task, the stimuli were presented in continuous alternation for one second each with no dark interval. The difference in variance between the two evaluation modes was significant ($p<0.05$) when observers controlled both the luminance and chromaticity of the test stimulus but was not significant when they had control of chromaticity only.

2.4 Summary

Simultaneous evaluations may suffer from positional bias, and there is evidence from previous studies of spatial brightness that counterbalancing can offset this bias [Fotios, Houser & Cheal, 2008]. Sequential evaluations may suffer from interval bias: this may be stronger in the successive evaluation where stimuli are seen only once than in the sequential evaluation where stimuli are seen repeatedly, but previous studies of spatial brightness do not present the null condition data which would quantify this. The data are inconclusive as to whether the precision of brightness judgements would differ for the two evaluation modes.

3. PHYSIOLOGICAL RESPONSE

When making a judgement between two stimuli of different SPD the evaluation mode will affect the SPD received at the eye. In simultaneous evaluations this is a mixed spectrum of the two stimuli, whereas in sequential evaluations it is the individual spectra in rapid succession. Two possible systems for mediating the SPD effect on spatial brightness are chromatic adaptation and pupil size.

3.1 Chromatic Adaptation

Spaces illuminated by lamps of different SPD can appear differently bright at the same illuminance because illuminance, as defined by The CIE Standard Photopic Observer (V_λ), is derived from a different visual process to that of brightness. The post-receptoral visual system is organized in three channels, one luminance channel where signals from the long- and medium-wavelength sensitive cone types are combined, and two colour channels where the differences between signals from different combinations of cone types are taken [Hunt, 1995]. The CIE Standard Photopic Observer is based on data collected primarily using flicker photometry and step-by-step brightness matching, techniques that tend to minimise activity in the colour channels; brightness perception is dependent on activity in all three channels [Lennie, Pokorny & Smith 1993, Wagner & Boynton 1972, Yaguchi & Ikeda, 1983].

Chromatic adaptation is the gradual neutralisation of activity in the opponent colour channels as the eyes acclimatise to the stimulus. Complete chromatic adaptation takes up to two minutes to achieve [Fairchild & Reniff, 1995; Shevell, 2001], and the observer's white point becomes the chromaticity of the stimulus, eliminating, or at least reducing, the brightness contribution from the opponent colour channels. In simultaneous evaluations the chromatic adaptation state of the observer is difficult to define. The observer does not adapt to the individual stimuli but to the mixed spectrum; Braun et al suggest "*an adapting white point … should be chosen somewhere between the two adapting conditions being considered*" [Braun, Fairchild & Alessi, 1996]. Studies of spatial brightness using sequential evaluation have presented the stimuli for up to five seconds each, showing both stimuli several times during each trial, and this would lead to approximately 60% chromatic adaptation [Fairchild & Reniff, 1995; Shevell, 2001]. The observer's white point would therefore swing between the chromaticities of the two stimuli, without reaching either.

Therefore, both simultaneous and sequential evaluations render incomplete chromatic adaptation. There will still be some activity in the opponent colour system and thus lamp SPD can affect brightness. Whether the two evaluation modes lead to different judgements of brightness may depend on the duration of exposure and the point during presentation within the sequential mode that decisions are made: if participants make a decision at the onset of each stimulus, then the adaptation to these stimuli will be minimal and the brightness difference large, but if the decision is made toward the end of the presentation period of each stimulus then there will be a greater degree of chromatic adaptation to that stimulus and brightness differences may be diminished. Chromatic adaptation in sequential evaluations in which each stimulus is presented for longer than five seconds will tend towards that of separate evaluations, in which the chromatic contribution to brightness is reduced relative to that of joint evaluations [Fotios, 2006].

3.2 Pupil Size

Two studies have proposed that spatial brightness is mediated by pupil size [Berman et al, 1990; Viénot et al, 2009]. While there is evidence that the pupil changes size in response to changes in illuminance and SPD there is no established mechanism to explain why a reduction in pupil size caused by changes in SPD should lead to the perception that brightness has increased. Viénot postulates an indirect mechanism might be in operation: *although less light reaches the retina* (when the pupil contracts) *the observer "feels" that the illumination is brighter* [Viénot et al, 2009].

The pupil reflex is driven by the intrinsically photosensitive retinal ganglion cell (ipRGC) [Koga & Takao, 2009]. The ipRGC has a peak sensitivity to short wavelength light [Koga & Takao, 2009] and thus as the short wavelength component of lighting increases, the signal from the ipRGC to the pupil increases. In one study this was attributed to CCT: stimuli of higher CCT were found to be brighter [Viénot et al, 2009] although other studies suggest that CCT does not correlate with brightness judgements [Boyce, 1977; Boyce & Cuttle, 1990]. In the second study the effect was characterised by the scotopic to photopic (S/P) luminance ratio [Berman et al, 1990]: the scotopic (rod) response also has peak sensitivity in the short wavelength region and therefore lighting of higher S/P ratio would appear brighter. For most white light sources, S/P ratios increase as CCTs increase.

Berman et al report pupil sizes (area) for a range of S/P ratios at two different illuminances (64 lux and 106 lux) [Berman et al, 1997] and these provide a guide as to the range of pupil size differences expected in brightness judgements. The S/P ratios of the two light sources were chosen to be high (C75 lamp) and low (WW lamp) and pupil sizes were measured during observation of one of four screens of different colours. The smallest pupil size (10.69mm^2; or 3.69mm diameter) was found with the C75 lamp at the higher illuminance when observing the white surface (S/P=2.25); the largest pupil size (19.13mm^2; or 4.94mm diameter) was found with the WW lamp at the lower illuminance when observing the reddish-brown surface (S/P=0.84).

Sequential and simultaneous evaluations of brightness will elicit different responses if the sequential evaluation allows sufficient time for a significant change in pupil size. The ipRGC responds to light slowly relative to the rapid response of rods and cones [Koga & Takao, 2009]. Previous judgements of spatial brightness have used exposures of three seconds [Vrabel et al, 1998] and five seconds [Berman et al, 1990] to each stimulus. Upon opening in the dark after adaptation to a large, bright (320 cd/m^2) field, the pupil can dilate from a diameter of 2.9mm to 4.7mm in three seconds and up to 5.6mm diameter in five seconds, [Wyszecki & Stiles, 1982]. Contraction on exposure of a dark adapted eye to the same bright field reduces the pupil from a diameter of 8.0mm to 3.3mm in three seconds, and reaches 3.1mm in 4.5 seconds. There may also be an initial delay (latency) in the response to stimulus onset of around 0.3 seconds [Kasthurirangan & Glasser, 2006; Feinberg & Podolak, 1965] and a momentary dilation due to the dark period between stimuli (darkness reflex) [Andreassi, 2000]. Thus it is reasonable to expect that the change in pupil size following a change of illuminant in brightness matching tests would be complete within three seconds.

For simultaneous evaluations the ipRGC is stimulated by the mixed SPD of the two light sources, and pupil size is modified to an arbitrary mid-way diameter between the diameters it would adopt if stimulated by the two sources separately. In this case, two stimuli of equal luminance but different SPD would appear equally bright because the pupil would have the same diameter for both. For sequential evaluations using three to five seconds duration per stimulus the pupil diameter has sufficient time to respond to the SPD of each stimulus, and thus the same two stimuli of equal luminance but different SPD would appear differently bright because the pupil would have a different diameter on exposure to each stimulus. What is not known is whether the change in pupil size is complete at the point within the stimulus observation at which the observer makes the judgement.

3.3 Summary

The chromatic and pupil size systems respond differently to sequential and simultaneous evaluations of brightness. If the chromatic system is the dominant system for spatial brightness then sequential and simultaneous evaluations will yield similar responses; if the pupil response is the dominant system then the two modes of evaluation will lead to different judgements of spatial brightness. Further research is needed to investigate the point within a sequential evaluation at which the brightness decision is made as this affects the degree of chromatic adaptation and completeness of the change in pupil size.

4. METHOD

The above discussions suggest that simultaneous and sequential evaluations of spatial brightness may yield different estimates of the magnitude of lamp spectrum effects and different levels of precision. Fotios & Cheal previously reported the results of brightness matching and brightness discrimination tests at mesopic levels, both using simultaneous evaluations [Fotios & Cheal, 2007]. This work was repeated using sequential evaluation to enable quantitative comparison. Direct comparison of simultaneous and sequential brightness discrimination has been carried out at photopic levels [Fotios & Houser, 2009] and will be reported in a further publication.

Previous simultaneous evaluations [Fotios & Cheal, 2007] used a pair of side-by-side booths, with separate light sources simultaneously illuminating each booth. Light was transported to the top of each booth through a light pipe, with an iris in the pipe to enable variation of illuminance. The sequential evaluations used only one of these booths (Figure 1). This was again observed from a distance of 1.0 metre which presented a visual field of approximately 37° high and 38° wide. Light from two different lamps was transported to the booth through separate light pipes, again using irises to adjust the illuminance. Placed in series with each iris was a leaf shutter that provided rapid on-off switching of the source without affecting the dimming adjustment. Luminance measurements show negligible differences in spatial distribution between lamps, between light from the two light pipes and between levels of dimming. The interior surfaces were painted matt grey (Munsell N5, r=0.2) and contained coloured objects, these being four pyramids 60mm high, one each made from red, green, yellow and blue card.

The two stimuli were presented in rapid succession: stimulus A for 5s; a dark interval of approximately 300ms; stimulus B (5s); dark interval (300ms); stimulus A (5s) etc. These durations were chosen to repeat the conditions used by Berman et al [Berman et al, 1990]. For the matching test this procedure was followed until the test participant was satisfied with their brightness match (typically three to eight repetitions). For the discrimination test the number of repetitions was limited to three, i.e. ABABAB. The on/off shutter (leaf shutter) was triggered manually with regard to a digital clock. The accuracy of the timing of this was determined using a video recording of the shutter and clock during a sample of trials. It was determined that the presentation duration of 5 seconds was accurate to within ±500ms.

Four lamps were used; a standard high pressure sodium (HPS), a compact fluorescent (CFL) and two types of metal halide (MH1, MH2), as defined in Table 1, these being the same lamps as used in previous work [Fotios & Cheal, 2007]. Using the HPS as the reference source gave four lamp combinations including a null condition. The order in which lamp pairs were presented was balanced between subjects.

In sequential brightness matching trials one of the two lamps in a pair was set by the experimenter to the reference illuminance. The test participant used the dimming control, a three-turn rotary dial, adjusting the illuminance of the second stimulus so that the two were matched, as-near-as-possible, for equal brightness. This procedure was carried out four times by each test participant to counterbalance dimming application (applied to both lamps in the pair) and dimming direction (starting from high and low initial illuminances). When the HPS lamp was used as the stimulus of fixed illuminance the reference illuminance was 7.5 lux, as measured at the centre of the floor of the booth. When the MH and CFL lamps were used as the stimulus of fixed illuminance the reference illuminance was 5.0 lux; this was expected to be approximately equally bright as the HPS at 7.5 lux and thus maintain a similar level of adaptation in both cases.

In sequential brightness discrimination trials, lighting from one lamp in each pair was set to the reference illuminance and lighting from the second lamp was set to a range of illuminances. At each presentation the test participant reported which stimulus appeared brighter, a forced choice task. This procedure was repeated by each test participant to counterbalance lamp nomination as reference or variable stimulus. When the HPS lamp in a pair was used as the stimulus of fixed illuminance, this being 7.5 lux, the CFL and MH lamps were presented at 2.0, 3.0, 5.0, 7.5, 10.0 lux. When the MH or CFL lamps in a pair was used as the stimulus of fixed illuminance, this being 5.0 lux, the HPS lamp was presented at 3.0, 5.0, 7.5, 10.0 and 15.0 lux. These ranges were chosen with expectation that the middle value would tend to appear the more equally bright as the fixed illuminance stimulus, thus avoiding a stimulus frequency bias [Fotios & Cheal, 2008], and are the illuminance steps of the S-series of lighting classes for residential roads [BS EN 13201: 2003]. With each participant, the HPS reference source remained in the same housing (iris, shutter & light pipe), but across the participant sample this was balanced between the two housings.

The HPS reference illuminance of 7.5 lux gave a mean luminance of 0.35 cd/m^2 on the back wall of the booth. The range of mean luminances on the back wall experienced in trials was from approximately 0.10 cd/m^2 at 2.0 lux to 0.70 cd/m^2 at 15.0 lux.

Twenty one naïve subjects were used, and these were paid for their participation. Thirteen were female and eight were male; nineteen subjects were aged between 18 and 44 years old and two were aged 45 to 54 years old. The size and breakdown of this sample is very similar to that used in the previous simultaneous matching trials [Fotios & Cheal, 2007]. Some subjects had been used in previous lighting experiments during which it was established that they had normal colour vision: for those subjects who had not previously participated in lighting experiments, normal colour vision was confirmed using the Farnsworth dichotomous D-15 test.

5 RESULTS

5.1 Brightness Matching Null Condition

An examination of the null condition data is used to validate the experiment. In null condition trials, HPS lamps were used in both intervals, one being the reference and was set to 7.5 lux. Four matches were made by each test participant, with the dimming control operating both lamps, each for two trials, starting from high and low initial illuminances. The two HPS lamps were nominally called the reference lamp (HPS), used as the reference for all trials, and the comparison lamp (HPSc), used as the second source for the null condition trials, but were otherwise identical. The mean illuminance ratio (HPSc/HPS) at equal brightness is 0.996 (std dev = 0.035, n=21), which is not a significant departure from unity (t-test).

For approximately half the trials these lamps were placed in the left-hand or right-hand housings behind the booth. When the HPS lamp used the right-hand housing the mean illuminance ratio (HPSc/HPS) is 1.005 (std dev = 0.030, n=10) and when it was placed in the left-hand housing the mean illuminance ratio (HPSc/HPS) is 0.987 (std dev = 0.039, n=11). The difference between the groups is not statistically significant (t-test) which suggests negligible difference between the lamp housings.

Previous work using simultaneous evaluations has revealed a conservative adjustment bias, in which the variable stimulus is set to a lower level than expected [Fotios, 2001; Fotios & Gado, 2005; Fotios, Houser & Cheal, 2008]. There is a common psychological tendency to adjust insufficiently in tasks that involve estimation via adjustment and it is manifest in a variety of sensory responses [LaBoeuf & Shafir, 2006]. To determine whether this bias was present in the current work the results are formatted as variable source and fixed source. The mean illuminance ratio (variable/fixed) is 0.992 (std dev = 0.039, n=21), which is not a significant departure from unity (t-test) and suggests that the conservative adjustment bias was not present in the sequential matching task.

Analysis of the results from the null condition trials carried out to validate the experimental procedure suggests that any bias in the brightness matching procedure was negligible.

5.2 Brightness Matching Between-Lamps

Results of the sequential brightness matching tests are shown in Table 2. The set of illuminance ratios within each lamp pair (including the null condition) were examined to determine whether they appeared to be drawn from a normally distributed population by examination of central tendency, dispersion, graphical representation and statistical analysis. It was concluded that the data are drawn from a normally distributed population and hence statistical analysis employed parametric tests. For each lamp pair, Table 3 shows that the mean illuminance ratio (test/HPS) is less than unity, and analysis using the t-test suggests that these differences are significant ($p<0.01$).

Table 2 also shows the results of the previous brightness matching study using simultaneous evaluations and the same set of lamps. [Fotios & Cheal, 2007]. In those trials, each of 21 test participants carried out the matching task four times for each lamp pair, counterbalancing the lamp to which dimming was applied and the initial illuminance of the variable stimulus. The trials were carried out at three reference illuminances (2.0, 7.5 and 15 lux); the results shown in Table 3 are those from the 7.5 lux reference illuminance, but the conclusions drawn are identical if the mean data across all three reference illuminances were used. Application of the t-test suggests the mean illuminance ratio in each lamp pair departs significantly from unity ($p<0.01$). There is little difference between mean illuminance ratios obtained using simultaneous and sequential matching.

The null condition data suggested no evidence of the conservative adjustment bias. The between-lamps data are broken down in Table 3 according to which lamp was varied by the test participant. Participants applied the dimming to both lamps in the pair on successive trials, and for each the trial commenced with the variable lamp set to both high and low initial illuminances. The results shown in Table 3 are derived from the mean of the results obtained with the high and low initial illuminances. The trend is for lighting from the variable stimulus to be set to a higher than average illuminance, an exaggerative response rather than a conservative response. However, this effect is significant for only one (CFL/HPS) of the three lamp pairs ($p<0.05$). Together with the null condition data, this suggests that the sequential matching task did not suffer from conservative adjustment bias as has been found in simultaneous matching tests. In these tests the application of dimming was counterbalanced, so the effect of any bias associated with the application of dimming, whether conservative or exaggerative, was averaged out in the results.

5.3 Precision in Brightness Matching

Comparison of the standard deviations for simultaneous and sequential evaluations in each lamp pair as reported in Table 2 does not suggest a difference in the precision between subjects. This same conclusion is drawn if the standard deviations are derived from all 84 trials for each lamp pair (21 subjects x 4 repeats).

However, there may be a difference in precision within subjects. Table 4 shows the mean within-subjects standard deviations: for each test participant the standard deviations of their four trials per lamp pair were determined and Table 5 shows the means of these. These data suggest a lower within-subjects standard deviation for the sequential evaluations than for simultaneous evaluations.

An individual test participant is better able to make the same setting on repeat trials with the sequential match, but the difference between subjects of this setting is not affected and both evaluations yield a similar mean response with the same precision.

5.4 Brightness Discrimination Null Condition

In null condition discrimination trials participants compared two identical lamps, the reference HPS lamp (HPS) and the comparison HPS lamp (HPSc). One lamp was set to provide an illuminance of 7.5 lux and the second to provide 3, 5, 7.5, 10 or 15 lux in random order. Both lamps were used as the fixed illuminance stimulus, and thus there are 42 discrimination judgements at each illuminance combination.

When the two stimuli provided different illuminances, the stimulus of higher illuminance was judged to be brighter for 100% of trials. When the two stimuli were of equal illuminance, 55% of the judgements reported that the HPSc lamp was brighter. According to analysis using Dunn Rankin variance stable rank sums [Dunn-Rankin et al, 2004] the difference between the HPS and HPSc lamps is not significant which suggests negligible bias.

5.5 Brightness Discrimination Between-Lamps

In these trials one lamp was presented at a fixed illuminance and the second lamp at five different illuminances, the middle illuminance predicted to be the least different in brightness with the fixed illuminance stimulus.

Where the CFL, MH1 or MH2 lamps were presented at equal or higher illuminance than the HPS lamp, then these were judged to be brighter for almost 100% of the trials (of these 252 comparisons, there was only one judgment that HPS lighting at 5 lux was brighter than MH1 lighting at 5 lux). When the CFL, MH1 or MH2 lamps were presented at an illuminance two or three steps of the S-series below that of the HPS lamp, then the HPS was judged to be brighter on almost 100% of the trials (of these 252 comparisons, again only one judgment corresponded to CFL lighting at 3 lux being brighter than HPS lighting at 7.5 lux).

Table 5 shows the results of the brightness discrimination trials when the HPS lamp was set to 7.5 lux and the CFL, MH1 or MH2 lamps were set to 5 lux. This condition was experienced twice by each test participant, once when the HPS lamp was fixed at 7.5 lux and once when the test lamp was fixed at 5.0 lux, and hence these results are derived from 42 observations. For the CFL/HPS and MH2/HPS the distribution of judgements for brighter stimulus were close to equally distributed for both stimuli. For the MH1/HPS however, the HPS was judged to be the brighter stimulus in the majority of trials. Analysis using Dunn Rankin variance stable rank sums [Dunn-Rankin et al, 2004] suggests that the difference between the two lamps is not significant in all three pairs. These conclusions replicate those drawn from the results of the discrimination test previously carried out using simultaneous evaluations [Fotios & Cheal, 2007].

Hence these results suggest that following a reduction in illuminance from 7.5 to 5.0 lux, one step of the S-series, a space lit by the CFL, MH1 and MH2 lamps will not appear different in brightness to a space lit by HPS lamps at 7.5 lux. This supports the illuminance reduction permitted when lighting in residential streets in the UK [BS5489-1:2003].

6 DISCUSSION

Table 6 shows the mean illuminance ratio for equal brightness under each lamp pair, for the tests using sequential evaluation as reported in this article, and for the tests using simultaneous evaluation as were previously reported [Fotios & Cheal, 2007]. For the brightness matching tests these are simply the mean illuminance ratios derived from the test results. For the brightness discrimination tests, the mean illuminance ratios for equal brightness were determined using the four-parameter logistic equation [Menon & Bhandarkar, 2004].

Two observations are drawn from Table 6. Firstly, there appears to be little difference in illuminance ratio for a particular lamp pair between sequential and simultaneous evaluation modes, for both the matching and discrimination tasks. Analysis using the two-tailed t-test to the brightness matching results does not suggest that differences between simultaneous and sequential matching tasks are significant. This suggests that the evaluation mode did not significantly affect operation of the visual mechanism responsible for spatial brightness at the mesopic illuminance at which these trials were conducted. Secondly, the brightness discrimination data suggest illuminance ratios that depart slightly further from unity than do those from the matching task: the one-tailed t-test suggests these differences are significant ($p<0.05$).

Table 6 suggests one possible difference, for the CFL/HPS lamp combination, where the illuminance ratio for the simultaneous discrimination task (0.59) is lower than for the other methods. This value was determined from the simultaneous discrimination trials carried out at 7.5 lux: if data from the trials carried out at 2.0 and 15 lux are included the CFL/HPS illuminance ratio would be 0.65, much closer to the other values. For the MH1/HPS and MH2/HPS lamp pairs there was little difference between illuminance ratios determined from any of the three reference illuminances used in the simultaneous discrimination tests.

Comparison of the standard deviations found in simultaneous and sequential matching (Table 2) suggest little difference in precision of the mean illuminance ratio determined using the two modes of evaluation. One advantage of the sequential matching task is that it does not appear to exhibit the conservative adjustment bias previously found in simultaneous matching. Advantages of the simultaneous matching task are that, anecdotally, test participants tended to prefer it, and it can be carried out in less time than sequential evaluations.

It was suggested above that if the chromatic system was the dominant system for judgements of spatial brightness then sequential and simultaneous evaluations of brightness would yield similar responses, but that if the pupil response was the dominant system then the two modes of evaluation would lead to different judgements of spatial brightness. The experimental results do not suggest a difference between the sequential and simultaneous evaluations of brightness and therefore that any change in pupil size has not significantly affected the brightness judgement. A similar conclusion was drawn from tests carried out at photopic levels of illumination [Houser, Fotios & Royer, 2009].

All subjective evaluations of a physical stimulus are biased; good experimental design should seek to remove the effect of bias from the recorded data. In continuation of previous work [Fotios, Houser & Cheal, 2008] this article has identified the need to counterbalance the location (left-right) or interval (first-second) in which stimuli are presented, the application of dimming control to both stimuli, and to commence the dimming process from high and low initial illuminances. Null condition trials should be employed to quantify the magnitude of bias. Multiple methods of comparing stimuli should be used: if each method points toward the same conclusion despite bias unique to each method, then more confidence can be placed in the conclusion.

7 CONCLUSION

The objective of this work was to compare brightness matching and discrimination judgements using sequential and simultaneous modes of evaluation. Comparison of the results gained from these tests suggests no difference in either the illuminance ratio required for equal brightness nor the precision of this estimate. Therefore, we suggest that both modes of evaluation have equal validity.

The test results provide further evidence that lamp SPD affects judgements of spatial brightness at mesopic levels of illumination. In the current work, brightness under the CFL, MH1 and MH2 sources was judged brighter than under HPS of equivalent illuminance: there was a difference in illuminance of approximately 70% between the CFL, MH1 and MH2 lamps and the HPS lamp for equal brightness.

Acknowledgement

This work was carried out with support from the Engineering and Physical Sciences Research Council (EPSRC) grant reference EP/F035624/1.

References

Andreassi JL. Psychophysiology: Human Behavior and Physiological Response. 4th Edition, 2000. Mahwah, NJ: Lawrence Erlbaum Associates

~~Berman SM, Correspondence: A new retinal photoreceptor should affect lighting practice. Lighting Research & Technology, 2008; 40(4); 373-376~~

Berman SM, Jewett DL, Fein G, Saika G & Ashford F. Photopic luminance does not always predict perceived room brightness, Lighting Research & Technology **22**(1) 37-41 (1990)

Berman SM, Jewett DL, Benson BR & Law TM. Despite different wall colours, vertical scotopic illuminance predicts pupil size. Journal of the Illuminating Engineering Society, 1997; 26(2) 59-68.

Boyce PR. Investigations of the subjective balance between illuminance and lamp colour properties, Lighting Research and Technology, 1977; 9; 11-24

Boyce PR & Cuttle C. Effect of correlated colour temperature on the perception of interiors and colour discrimination, Lighting Research & Technology, 1990; 22(1); 19-36

Braun KM, Fairchild MD and Alessi PJ. Viewing techniques for cross-media image comparisons, Color Research & Application, 1996; 21(1); 6-17

British Standards Institution (BSI) BS EN 13201-2:2003, Road lighting - Part 2: Performance requirements, London: BSI, 2003

British Standards Institution (BSI) BS 5489 Road Lighting: Part 3, Code of Practice for Lighting for Subsidiary Roads and Associated Pedestrian Areas, London: BSI, 1992.

Bullough JD, Yuan Z & Rea MS. Perceived brightness of incandescent and LED aviation signal lights, Aviation, Space and Environmental Medicine, 2007; 78(9); 893-900

Dunn-Rankin P, Knezek GA, Wallace S & Zhang S. Scaling methods. 2nd Edition, 2004; Lawrence Erlbaum Associates, Mahwah, New Jersey

Fairchild MD & Reniff L. Time course of chromatic adaptation for color-appearance judgements, Journal of the Optical Society of America, series A, 1995; 12(5): 824-833

Feinberg R & Podolak E. Latency of the pupillary reflex to light stimulation and its relationship to aging. Washington DC: Federal Aviation Agency, Office of Aviation Medicine, 1965. Reprinted from Behavior, aging and the nervous system, Welford AT, Birren JE (ed.s), Springfield, Illinois; Charles C Thomas.

Foster DH, Amano K & Nascimento MC. Colour constancy from temporal cues: better matches with less variability under fast illuminant changes, Vision Research, 2001; 41: 285-293

~~Fotios SA, Lamp colour properties and apparent brightness: a review. Lighting Research and Technology, 2001; 33(3); 163-181.~~

Fotios SA. An error in brightness matching associated with the application of dimming, Lighting Research & Technology, 2001; 33(4); 223-231

Fotios SA. The Effect Of Chromatic Adaptation Upon The Relationship Between Lamp Spectrum And Brightness At Photopic Levels, Lighting Research & Technology, 2006; 38(1); 3-17

Fotios SA & Cheal C. Lighting in subsidiary streets: investigation of SPD effects. Part 2 – Brightness, Lighting Research & Technology, 2007; 39(3); 233-252

Fotios SA & Cheal C. Evidence For Response Contraction Bias In Side-By-Side Matching Tasks, Lighting Research & Technology, 2007; 39(2); 159-169

Fotios SA & Cheal C. The Effect of a Stimulus Frequency Bias in Side-By-Side Brightness Ranking Tests, Lighting Research & Technology, 2008; 40(1); 43-54

Fotios S & Gado T. A comparison of visual objectives used in side-by-side matching tests, Lighting Research & Technology, 2005; 37(2); 117-131

Fotios SA & Houser K. Tuning the spectrum of lighting to enhance spatial brightness: Investigations of research methods, Experiencing Light, Eindhoven, 26-27 October 2009,

Fotios SA, Houser KW & Cheal C. Counterbalancing needed to avoid bias in side-by-side brightness matching tasks, Leukos, 2008; 4(4); 207-223

Fotios SA & Levermore GJ. The Perception Of Electric Light Sources Of Different Colour Properties, Lighting Research & Technology, 1997; 29(3); 161-171

Hsee CK, Loewenstein GF, Blount S & Bazerman MH. Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. Psychological Bulletin, 1999; 125(5); 576-590

Hunt RWG. Measuring Colour. Second edition. 1995. Ellis Horwood, London

Jäkel F & Wichmann FA. Spatial four-alternative forced-choice method is the preferred psychophysical method for naïve observers, Journal of Vision, 2006; 6; 1307-1322

Kinney JAS. Sensitivity of the eye to spectral radiation at scotopic and mesopic intensity levels. Journal of the Optical Society of America, 1955; 45(7); 507-514

LaBoeuf RA & Shafir E. The long and short of it: Physical anchoring effects, Journal of Behavioural Decision Making, 2006; 19; 393-406

Kasthurirangan S & Glasser A. Age related changes in the characteristics of the near pupil response. Vision Research, 2006; 46; 1393-1403.

Koga Y & Takao M. Light response properties of intrinsically photosensitive retinal ganglion cells in non-image-forming vision. CIE Light & Lighting Conference, 27-29 May 2009, Budapest. Programme and abstracts, 145-146

Lennie P, Pokorny J & Smith VC. Luminance. Journal of the Optical Society of America (A), 1993; 10(6): 1283-1293

Menon A & Bhandarkar S. Predicting polymorphic transformation curves using a logistic equation. International Journal of Pharmaceutics, 2004; 286; 125–129

~~McNelis JF, Howley JG, Dore GE & DeLaney WB, Subjective appraisal of colored scenes under various fluorescent lamp colors, Lighting Design & Application, **15**(6) 25-29 (1985)~~

~~Murray A, Lawrence GP, Clayton RH. Repeatability of dynamic eye pupil response measurement using the Pupilscan instrument. Clinical Physics and Physiological Measurement, 1991; 12(4); 377-385.~~

Shevell SK. The time course of chromatic adaptation. Color Research & Application, 2001; Supplement 26: S170-S173

Teller DY, Pereverzeva M & Civan AL. Adult brightness vs. luminance as models of infant photometry. Journal of Vision, 2003; 3; 333–346

Uchikawa K & Ikeda M. Accuracy of memory for brightness of colored lights measured with successive comparison method, Journal of the Optical Society of America A, 1986; 3; 34-39

Viénot F, Durand M-L, Mahler E. The effect of LED lighting on performance, appearance and sensations. CIE Light & Lighting Conference, 27-29 May 2009, Budapest. Programme and abstracts, 19-20

Vrabel PL, Bernecker CA & Mistrick RG. Visual performance & visual clarity under electric light sources: Part II - Visual Clarity. Journal of the Illuminating Engineering Society, 1998; 27(1); 29-41

Wagner G & Boynton RM. Comparison of four methods of heterochromatic photometry, Journal of the Optical Society of America, 1972; 62(12); 1508-1515

Wyszecki G & Stiles WS. Colour science: Concepts and methods, quantitative data and formulae. 2nd edition, 1982. New York: John Wiley & Sons Inc.

Yaguchi H & Ikeda M. Contribution of opponent-colour channels to brightness. in Mollon JD, Sharpe LT (Eds), Colour Vision: Physiology & Psychophysics. Academic Press Inc. (London) Ltd; London. 1983. 353-360

Yeshurun Y, Carrasco M & Maloney LT. Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. Vision Research, 2008; 48; 1837-1851

Zheleznikova OE & Myasoedova EI. Comparison of luminosities of lighting installations with discharge light sources, Light & Engineering, 1995; 3(2); 38-39

Lamp		**CRI**	**CCT (K)**	**S/P**
HPS, HPSc	*70W SON-T Pro*	25	2000	0.51
CFL	*55W PL-L*	82	3000	1.42
MH1	*70W CDO-TT*	83	2800	1.13
MH2	*70W CDM-T*	92	4200	1.65

Table 1. Lamps used in sequential brightness judgements. These are the lamps as used by Fotios & Cheal in previous brightness matching and discrimination tests [Fotios & Cheal, 2007]. S/P ratios calculated from SPDs measured in the booth: CCT and CRI as reported in lamp manufacturer's literature.

	Lamp combination		
	CFL/ HPS	**MH1/ HPS**	**MH2/ HPS**
Sequential matching			
mean illuminance ratio	0.704	0.752	0.710
std dev	0.085	0.080	0.089
n	21	21	21
departure from unity (t-test)	p<0.01	p<0.01	p<0.01
Simultaneous matching			
mean illuminance ratio	0.718	0.733	0.724
std dev	0.070	0.091	0.086
n	21	21	21
departure from unity (t-test)	p<0.01	p<0.01	p<0.01

Table 2. Results of brightness matching tests; mean illuminance ratio at equal brightness. Results of simultaneous matching task as previously reported [Fotios & Cheal, 2007].

Variable stimulus		**Lamp pair (test/HPS)**		
		CFL/HPS	**MH1/HPS**	**MH2/HPS**
Test lamp (CFL, MH1, MH2)	Mean illuminance ratio (test/HPS)	0.731	0.760	0.732
	std dev	0.112	0.099	0.123
	n	21	21	21
HPS lamp	Mean illuminance ratio (test/HPS)	0.677	0.744	0.689
	std dev	0.082	0.102	0.100
	n	21	21	21
Difference (t-test)		p<0.05	n.s.	n.s.

Table 3. Results of brightness matching trials broken down according to the application of dimming. (n.s. = not statistically significant.)

	Mean within-subjects standard deviation			
	HPSc/HPS	**CFL/HPS**	**MH1/HPS**	**MH2/HPS**
Simultaneous evaluations [Fotios & Cheal, 2007]	0.078	0.088	0.117	0.123
Sequential evaluations (current work)	0.070	0.073	0.087	0.092

Table 4. Mean within-subjects standard deviations for sequential and simultaneous brightness matching tests. (Note: for HPSc/HPS simultaneous, 18 subjects carried out 8 trials; for all other cases, 21 subjects carried out 4 trials).

Lamp pair	**CFL/HPS**	**MH1/HPS**	**MH2/HPS**
Percentage of votes for HPS brighter	55%	74%	52%
Percentage of votes for test lamp brighter	45%	26%	48%
Difference between lamp brightnesses	n.s.	n.s.	n.s.

Table 5. Results of brightness discrimination tests when the illuminances were 5.0 lux for the test lamp and 7.5 lux for the HPS lamp. Difference between lamps tested using Dunn Rankin variance stable rank sums: n.s. = not significant.

Response method	**Evaluation mode**	**Illuminance ratio at equal brightness**			
		HPS/ HPS	**CFL/ HPS**	**MH1/ HPS**	**MH2/ HPS**
Brightness matching	Sequential	1.00	0.70	0.75	0.71
	Simultaneous	0.99	0.72	0.73	0.72
Brightness discrimination	Sequential	0.99	0.67	0.69	0.67
	Simultaneous	1.00	0.59	0.68	0.64

Table 6. Comparison of illuminance ratios for equal brightness determined using matching and discrimination tasks with simultaneous and sequential modes of evaluation. Simultaneous data as previously reported [Fotios & Cheal, 2007]; these values are derived from the tests carried out at 7.5 lux. All values rounded to two decimal places.

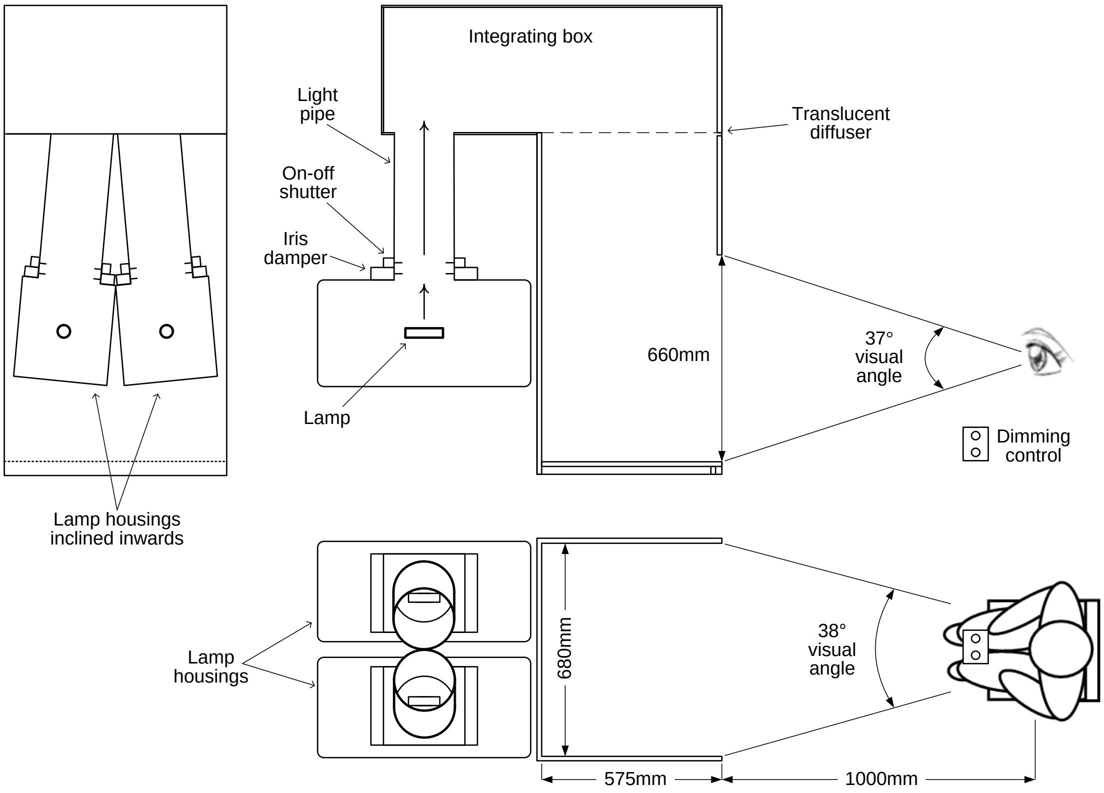

Figure 1. Diagram of the test apparatus. The two light pipes are aimed towards the same point in the roof of the integrating box.