Colour Constancy as Measured by Least Dissimilar Matching

Alexander D. Logvinenko * and Rumi Tokunaga 1
Department of Vision Sciences, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, UK
1 Present address: Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira, Aoba, Sendai 980-8577, Japan

Received 17 March 2011; accepted 23 June 2011

Abstract
Although asymmetric colour matching has been widely used in experiments on colour constancy, an exact colour match between objects lit by different chromatic lights is impossible to achieve. We used a modification of this technique, instructing our observers to establish the least dissimilar pair of differently illuminated coloured papers. The stimulus display consisted of two identical sets of 22 Munsell papers illuminated independently by neutral, yellow, blue, green and red lights. The lights produced approximately the same illuminance. Four trichromatic observers participated in the experiment. The proportion of exact matches was evaluated. When both sets of papers were lit by the same light, the exact match rate was 0.92, 0.93, 0.84, 0.78 and 0.76 for the neutral, yellow, blue, green and red lights, respectively. When one illumination was neutral and the other chromatic, the exact match rate was 0.80, 0.40, 0.56 and 0.32 for the yellow, blue, green and red lights, respectively. When both lights were chromatic, the exact match rate was found to be even poorer (0.30 on average). Yet, least dissimilar matching was found to be rather systematic. Particularly, a statistical test showed it was symmetric and transitive. The exact match rate was found to be different for different papers, varying from 0.99 (black paper) to 0.12 (purple paper). Such a variation can hardly be expected if observers’ judgements were based on an illuminant estimate. We argue that colour constancy cannot be achieved for all the reflecting objects because of mismatching of metamers. We conjecture that the visual system might have evolved to have colour constant perception for some ecologically valid objects at a cost of colour inconstancy for other types of objects.

Keywords
Colour, colour vision, colour constancy, dissimilarity judgment, mismatching of metamers, object colour

* To whom correspondence should be addressed. E-mail: a.logvinenko@gcu.ac.uk

© Koninklijke Brill NV, Leiden, 2011

DOI:10.1163/187847511X588746
1. Introduction

Object colour is generally believed to remain relatively constant when the object illumination changes. A textbook example is a banana looking yellow under different phases of daylight despite the fact that the spectral power distribution of the light significantly changes. This phenomenon has been conceptualised in the notion of colour constancy with respect to illumination (Brainard, 2004; Foster, 2011; Jameson and Hurvich, 1989; Katz, 1911/1999; Pokorny et al., 1991; Shevell and Kingdom, 2008). The deviations from constancy registered in experiments are usually thought to be insufficient to undermine this belief (Brainard, 2009b; Hurlbert, 2007; Smithson, 2005).

Still, from everyday experience we also know that objects might change their colour appearance due to illumination. Another textbook example is the colour of a dress which can be rather different indoors and outdoors. Such an illuminant dependence of object colour has been conceptualised in the notion of colour rendering (Logvinenko, 2009a; Schanda, 2007; Wyszecki and Stiles, 1982). The more an illuminant affects the colour of objects, the poorer its colour rendering.

Although the notions of colour rendering and colour constancy seem to be mutually exclusive, they both capture important aspects of the effect which illumination has on the colour of objects. Indeed, on the one hand the colour of an object apparently changes when its illumination changes. If full colour constancy existed then we would not be able to distinguish areas lit by different lights. However, observers are known to readily discriminate between illumination and material borders in variegated scenes with multiple light sources (Craven and Foster, 1992; Foster, 2008; Foster et al., 2001). On the other hand, the more the colour of the object changes due to illumination, the more it becomes obvious that some important aspects of the colour remain the same. What in the colour of objects changes when the illuminant changes, and what remains constant then?

One, rather old view is that we are capable of perceiving separately and simultaneously the colour of objects and the colour of illumination (Helmholtz, 1867; MacLeod, 2003; Mausfeld, 2003). Speaking more formally, object colour and illumination colour are assumed to be separable and independent attributes, similar, for example, to the width and height of rectangles. Therefore, when the illumination changes we see a change in the illumination colour, with the object colour remaining the same.

Note, firstly, that this theory implies the existence of a single palette of object colours (see Note 1) which does not depend on illumination. However, there is evidence that the object colour palette changes with illumination. Particularly, the palette of achromatic object colours was shown to shrink when the illumination intensity is decreased (Logvinenko and Maloney, 2006).

Secondly, this theory implies the possibility of asymmetric (across illuminations) colour matching. Indeed, the same way as one can match the width of rectangles of different height, one should be possible to make a match with respect to object colour despite the difference in illumination colour. However, there is abundant
evidence that asymmetric colour match is hard to achieve (Brainard et al., 1997; Logvinenko and Maloney, 2006).

We argue elsewhere that asymmetric colour matching is impossible because alongside the classical three dimensions of object colour, which mainly correlate with the material (reflectance) properties of an object, there exist three additional dimensions of object colour which mainly correlate with the object’s illumination (Tokunaga and Logvinenko, 2010b). We refer to these as material and lighting colour dimensions of object colour. It should be stressed that although lighting colour dimensions relate to illumination, they are not the dimensions of illumination colour because they are not solely determined by the illuminant. An alternative way to express this fact is to assert that the object colour palette varies with illumination. In other words, for different illuminations there exist different object colour palettes.

Thus, in a multiple light scene one can differentiate colours within the same object colour palette (the material colour difference) and across the different object colour palettes (the lighting colour difference). Note that the differences of both sorts are understood as differences in object colour. The existence of unavoidable lighting colour differences in a scene with at least two sources of illumination makes an exact asymmetric colour match impossible.

An intriguing fact is that while an exact asymmetric colour match is impossible, observers have proved capable of establishing approximate asymmetric colour matches. It is this ability that justifies the colour constancy metaphor. An obvious question arises then. What do observers actually do when they are asked to make an asymmetric colour match if an exact match is impossible? One hypothetical answer might be that they perform a least-dissimilar match. In other words, having found out that there is no exact match for a test paper, observers might try to find a paper least dissimilar from the test paper.

To test this hypothesis Logvinenko and Maloney (2006) explicitly asked their observers to find the least dissimilar pair between two identical series of achromatic Munsell papers illuminated by different lights. It was found that the papers lit by different lights were chosen as least dissimilar when their albedos were practically the same (Logvinenko and Maloney, 2006). This result is in line with observations made in subsequent experiments on multidimensional scaling of Munsell papers illuminated by different lights (Tokunaga and Logvinenko, 2010a, b, c; Tokunaga et al., 2008). In these experiments observers were asked to estimate subjective dissimilarity between the papers. It was found that on average the least dissimilarity was assigned to the physically identical papers (lit by different lights).

These results lead to the hypothesis that when comparing two differently illuminated objects, the dissimilarity between them achieves its minimum when these have the same material colour. The idea behind this hypothesis is that, given some fixed difference in lighting colour dimensions, introducing an additional difference in material colour dimensions can only increase dissimilarity.
Therefore, using the least dissimilar task one can ascertain, at least to a first approximation, which papers have the same material colour under different illuminations. That is, the least dissimilar task can be an alternative to the asymmetric colour matching task.

In this article we report on an experiment that follows up on a previous study (Logvinenko and Maloney, 2006) by extending it into the chromatic domain. As in the previous study we used real reflecting objects (papers) lit by real lights rather than their computer simulation. Although one can produce quite realistic images on the monitor screen these cannot always be a satisfactory substitute for real objects in experiments on colour vision. As a matter of fact, the colour perception of a real object and its pictorial reproduction are not identical. In particular, a three-dimensional paper implementation of Adelson’s picture of the wall-of-blocks inducing a strong lightness illusion (Adelson, 1993), was found to bring about no illusion at all (Logvinenko et al., 2002). That is, the lightness perception of the three-dimensional material prototype of the wall-of-blocks was found to be veridical despite the fact that both the real thing and its picture produced practically equal luminance patterns in the eye. Another striking example of the difference in colour perception between real scenes and their pictorial reproduction, especially pertinent in the present context, was described by Logvinenko et al. (2008) who performed a multidimensional scaling analysis of Adelson’s snake-like pattern (Adelson, 2000). Although this picture renders a strong difference in apparent illumination between its different areas, multidimensional analysis showed no traces of the second dimension (surface-brightness) of achromatic colour (Logvinenko et al., 2008) which revealed itself in a previous study employing material papers and real lights (Logvinenko and Maloney, 2006). In other words, the snake-like picture is perceived as a lightness pattern the surface-brightness of which is homogeneous (that is, actually, in line with the physical nature of the stimulus: a sheet of paper contaminated with ink). All these results convince us that if our goal is to study the real (not pictorial) colour perception we have to use real (not pictorial) stimuli.

More specifically, we conducted an experiment using an experimental design typical for colour constancy studies but with a different instruction. We asked our observers to perform a dissimilarity judgment rather than make an asymmetric colour match. While the dissimilarity judgment task is more flexible than the exact match task, both involve manipulation of the object colours. As one cannot gradually adjust reflecting properties of an object, the only alternative is to use a discrete set of objects large enough to secure accurate judgements (see Note 2).

2. Experimental

Four normal trichromats took part in the experiment. All of them had participated in colour experiments before. Except for Obs. 1 (the coauthor RT), they were naive with respect to the purpose of the experiment.

The stimulus display, consisting of two fields made of sheets of white paper, was mounted on a wall covered by black cloth. The sheets of paper, $53 \times 74$ cm
Figure 1. A sketch of the experimental display. It illustrates the lighting condition when the left field is lit by the yellow light, and the right field by the blue light. This figure is published in color in the online version.

Each, were 3.5 cm apart. Two random configurations of the same set of Munsell papers were placed in these fields (Fig. 1). Each stimulus array contained 20 papers from every other page in the Munsell book of maximal Chroma (Fig. 2) completed with grey (N5/) and black (N1/) papers (i.e., 22 papers in all). The stimulus array dimensions were 39 × 55 cm.

Six different lights were used to independently illuminate each field of the display: neutral (N), yellow (Y), blue (B), green (G) and two reds (R1 and R2). The lights were produced by Kodak projectors equipped with chromatic filters. The spectral power distributions of these lights are depicted in Fig. 3. Figure 4 presents the CIE 1931 chromaticity coordinates of the stimulus papers under all the six lights. All the six lights produced approximately equal illuminance (128–135 lux).

The experimental procedure was as follows. Having pointed out with a laser pointer a paper in one-half of the display (referred to as the test), the experimenter asked the observer to find a paper in the other half that was least dissimilar to the test. Although it is rather loose usage of the term, we will call the observer’s choice the match.

The experiment was divided into sessions. The illumination of the fields was kept constant during a session. All the possible 6 × 6 combinations of lights (referred to as illumination conditions) were exhausted. We will use notation Y–N, or R1–G and so on for illumination conditions. Specifically, R1–G means that the left field was
Figure 2. Photograph of the chromatic stimulus papers used in the experiment. Their Munsell notations are: (1) 5RP5/12; (2) 10P4/12; (3) 5P4/12; (4) 10PB4/12; (5) 5PB5/12; (6) 10B5/12; (7) 5B5/10; (8) 10BG5/10; (9) 5BG6/10; (10) 10G5/10; (11) 5G5/10; (12) 10GY6/12; (13) 5GY7/12; (14) 10Y8.5/12; (15) 5Y8/14; (16) 10YR7/14; (17) 5YR7/14; (18) 10R5/16; (19) 5R4/14; (20) 10RP5/14. This figure is published in color in the online version.

Figure 3. Spectral power distribution of the illuminants employed in the experiment. The line colour matches the colour of the light each line represents. The red #2 light is represented by the purple line (with the rightmost peak and no short-wavelength ‘lobe’). This figure is published in color in the online version.

illuminated by the red #1 light, the right field by the green light. During a session, the papers in one field served as test, in the other field as match. Each paper in the test field was pointed out in random order by the experimenter only once. The Munsell papers in a stimulus array were arranged randomly and differently in each session. Each field illuminated by a particular light was used three times as the test field and three times as the match field. Thus, each session for every one of the 36 illumination conditions was repeated three times for each observer.

The experimental room was semi-darkened. Observers sat at a distance of 1.5 m from the experimental display. They saw the projectors casting light on the display,
so observers were fully aware of the nature of the stimuli: real papers lit by real lights. They were adapted to each illumination condition for 10 min before each experimental session. Vision was binocular and free of any restrictions. Observers were advised to move their gaze between the left and right fields in order to minimise local chromatic adaptation. A session took 15–20 min.

3. Results

Figures 5–11 show all the matches made by one observer (Obs. 2). The results of the other three observers are rather similar. Figure 5 presents the results obtained under symmetric illumination conditions, that is, when the test and match papers were lit by the same light. Points on the diagonal indicate exact matches, that is, the cases when the observer pointed out in the match field the same paper as that indicated by the experimenter in the test field. Note that all three matches are presented in each plate. When there is only one marker on the diagonal it means that all the three matches were the same for this test paper. The percentages of exact matches (referred to as the exact match rate) made by all the observers when both lights were identical are given in Table 1. The exact match rates in this table reflect observers’ accuracy of matching under symmetric illumination conditions. On average the most accurate matches were made under the yellow (93%) and neutral (92%) illuminations. However, even under these lights some mismatches were made by all observers, especially in the blue–green part of the hue circle. Matching was the worst under the red lights. On average, nearly a quarter of all matches were incorrect for the red illuminations (76 and 75% for red #1 and red #2, respectively).
Figure 5. Least dissimilar matches made by one observer when the test and match fields were illuminated by the same light. The test and match papers, numbered as in Fig. 2, are plotted on the horizontal and vertical axes, respectively. The marker colour indicates the colour of the match paper. The physically exact matches lie on the diagonal. The letters in the brackets in the axes labels designate the illuminants. This figure is published in color in the online version.

Figure 6 presents all the matches made by Obs. 2 when the test illuminant was neutral. Figures 7–11 show the matches of the same observer for the other five test illuminants. This observer’s results are summarised in Table 3, where the exact match rates are presented for all the 36 illumination conditions. Tables 2, 4 and 5 show the exact match rates for the rest of three observers. The averaged exact match rates are given in Table 6.

Close inspection of Fig. 6 shows that the matches under the N–Y illumination condition were nearly as good as under the N–N illumination condition (i.e., symmetric matches). Indeed, very few points in the middle plate in the upper row in Fig. 6 deviate from the diagonal by more than one hue step (see Note 3). (Note that as the chromatic stimulus papers make a hue circle, some apparently large deviations from the diagonal — such as the three matches in the upper left corner of the plate corresponding to the N–B illumination condition in Fig. 6 — are, in fact, much closer to the test papers.) Therefore, although the exact match rate for the N–Y illumination condition is lower than that for the N–N illumination condition, the degree of mismatching for the N–Y illumination condition is not much worse.
Figure 6. Least dissimilar matches made by one observer when the test field was illuminated by the neutral light. The match field illumination is marked by the letter in the brackets in the vertical axis labels. The test and match papers, numbered as in Fig. 2, are plotted on the horizontal and vertical axes, respectively. The marker colour indicates the match paper colour. The physically exact matches lie on the diagonal. This figure is published in color in the online version.

Averaging the data over observers, match illuminants, and chromatic papers, we have evaluated the mean mismatch rate (in hue steps) for each test illuminant (Table 7). As may be seen, when the test illuminant was neutral or yellow the average mismatch was roughly one hue step. The mismatch for the other four test illuminants was approximately two hue steps. Therefore, while the exact match rate for these illuminations (Table 8) is quite low (less than 30%) the average mismatch does not exceed two hue steps.

When performing these calculations we excluded those matches in which either of the two achromatic papers (#21 or #22) was chosen by the observer (referred to as achromatic mismatches for chromatic papers). The achromatic mismatch rate for each test illuminant is given (in percentage) in the first column of Table 8.

The exact match rate for chromatic papers for each test illuminant is presented in the second column of Table 8. It is nearly 40% of that for the neutral and yellow illuminations, the worst rate being for the blue light (nearly 20%). On average, the exact match rate for chromatic papers for all the illumination conditions was approximately 30%. In the third column of this table is the percentage of matches for which the mismatch was not more than one hue step, that is, either a zero (i.e.,
Figure 7. Least dissimilar matches made by one observer when the test field was illuminated by the yellow light. The match field illumination is marked by the letter in the brackets in the vertical axis labels. The test and match papers, numbered as in Fig. 2, are plotted on the horizontal and vertical axes, respectively. The marker colour indicates the match paper colour. The physically exact matches lie on the diagonal. This figure is published in color in the online version.

As we can see, on average (the bottom row in Table 8) the mismatch between chromatic test and match papers did not exceed 2 hue steps in more than 70% trials. Mismatch of more than 3 hue steps occurred in only 20% of the trials. Even for the blue illumination, for which the exact match rate was lowest, the mismatch exceeded 3 hue steps in only 25% of the trials. Note that the indices in all the columns but the first one in Table 8 are very well correlated. Therefore, each of them, including the exact match rate, can be rather informative as an index of the asymmetric colour matching.

Interestingly, the mean mismatch was found to be quite different for different papers (Fig. 12). It varies from less than one hue step for yellowish papers (5GY7/12, 10Y8.5/12, and 5Y8/14) to more than three hue steps for the purple paper (10RP5/14). Similar differences between papers have been also found in terms of the match rates evaluated separately for each paper (Fig. 13). The exact match rate curve (pentagrams in Fig. 13) seems to have the inverted shape of the curve in
Figure 8. Least dissimilar matches made by one observer when the test field was illuminated by the blue light. The match field illumination is marked by the letter in the brackets in the vertical axis labels. The test and match papers, numbered as in Fig. 2, are plotted on the horizontal and vertical axes, respectively. The marker colour indicates the match paper colour. The physically exact matches lie on the diagonal. This figure is published in color in the online version.

Fig. 12. The yellowish papers show higher colour constancy in terms of all the four indices of match rate represented in Fig. 13.

The achromatic papers prove to have an exact match rate that is higher than the chromatic ones. The perception of the black paper was remarkably robust: only 5 mismatches in 432 trials (i.e., 1.2%). In other words, under all the chromatic illuminations the black paper was found as subjectively least dissimilar. The mismatch rate for the grey paper was higher: 29%. Still, the grey paper has been found to be more constant than all the chromatic papers.

It follows from Table 8 that there is a considerable difference between the neutral and yellow test lights on the one hand, and the other test lights on the other. The mismatch rate for the blue, green, and both red test illuminations is considerably higher. In Figs 7–11 there are a few deviations from the diagonal by 5 and even more hue steps for these illuminations. For some illumination conditions, some papers happen to be chosen as a match for 5 or 6 adjacent test papers. Interestingly, observers were not aware that they chose systematically the same paper as a match for distinctively different test papers, and were very surprised when the experimenter informed them of that after the experiment was over.
Figure 9. Least dissimilar matches made by one observer when the test field was illuminated by the green light. The match field illumination is marked by the letter in the brackets in the vertical axis labels. The test and match papers, numbered as in Fig. 2, are plotted on the horizontal and vertical axes, respectively. The marker colour indicates the match paper colour. The physically exact matches lie on the diagonal. This figure is published in color in the online version.

Under all the test lights a certain proportion of times the neutral papers (predominately, the grey one) were chosen as a match for chromatic test papers (see achromatic mismatch rate in Table 8). Such achromatic mismatches occurred when some papers, for example, bluish papers under the red illuminations, looked virtually achromatic, losing nearly all their chromatic content.

Generally, it must be said that Munsell papers under strong chromatic illumination, as in our experiment, change their appearance rather drastically. Although they retain the circular chromaticity structure, the chromaticity difference between adjacent papers becomes considerably smaller, and for some papers (e.g., blue–green) disappears entirely. At the same time, the achromatic component increases. All this makes the matching task rather difficult. The typical problem with matching as it appeared to Obs. 1 (the coauthor RT) can be formulated as follows. “I realise that the test paper belongs to, say, the blue–green segment of the hue circle. However, none of the blue–green papers in the matching field looks similar enough to make a confident match.”

Finally, it should be mentioned that according to the report of Obs. 1, although she was fully aware of the difference in illumination between the fields, there was
no doubt that it was object colour that underwent the apparent alterations described above. In other words, when illumination becomes, say, red, a blue–green paper does not retain any appearance of blue–green. It becomes a purplish paper — a purplish paper observed ‘through a veil’ of red illumination. In other words, a change of the illuminant chromaticity results in not only a change in lighting colour, but also a change in material colour. That the difference in lighting colour did not hamper matching in material colour is supported by the high exact match rate for the grey paper: 71%. Moreover, as pointed out above, despite the presence of strong chromatic illumination, observers did not confuse the black paper with any other, not even with the grey paper at all.

4. Analysis of the Results

There is a general belief that asymmetric colour matching can be accounted for by either appropriate normalisation of the cone responses or differencing them with subsequent normalisation (Brainard, 2009b; Foster, 2008; Mausfeld, 1998; Smithson, 2005). In other words, colour is believed to be determined by cone contrast
Figure 11. Least dissimilar matches made by one observer when the test field was illuminated by the red #2 light. The match field illumination is marked by the letter in the brackets in the vertical axis labels. The test and match papers, numbered as in Fig. 2, are plotted on the horizontal and vertical axes, respectively. The marker colour indicates the match paper colour. The physically exact matches lie on the diagonal. This figure is published in color in the online version.

Table 1.
Exact match rates for symmetric colour matching

<table>
<thead>
<tr>
<th>Illumination condition</th>
<th>Obs 1</th>
<th>Obs 2</th>
<th>Obs 3</th>
<th>Obs 4</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>N–N</td>
<td>97</td>
<td>94</td>
<td>82</td>
<td>94</td>
<td>92</td>
</tr>
<tr>
<td>Y–Y</td>
<td>89</td>
<td>95</td>
<td>92</td>
<td>95</td>
<td>93</td>
</tr>
<tr>
<td>B–B</td>
<td>82</td>
<td>91</td>
<td>76</td>
<td>88</td>
<td>84</td>
</tr>
<tr>
<td>G–G</td>
<td>79</td>
<td>77</td>
<td>74</td>
<td>80</td>
<td>78</td>
</tr>
<tr>
<td>R1–R1</td>
<td>76</td>
<td>73</td>
<td>76</td>
<td>80</td>
<td>76</td>
</tr>
<tr>
<td>R2–R2</td>
<td>79</td>
<td>70</td>
<td>73</td>
<td>80</td>
<td>75</td>
</tr>
</tbody>
</table>

rather than cone response directly. Arguing in support of such a view, Paul Whittle even coined a term ‘contrast colours’ (Whittle, 2003). There are a few models of colour appearance based on cone differencing and normalisation (for a review see Hurlbert, 1998; Shevell, 2003; Whittle, 2003). We tested the predictions of some of them against our results.
Table 2.
Exact match rates for Obs. 1

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Y</th>
<th>B</th>
<th>G</th>
<th>R1</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>97</td>
<td>80</td>
<td>47</td>
<td>47</td>
<td>33</td>
<td>29</td>
</tr>
<tr>
<td>Y</td>
<td>83</td>
<td>89</td>
<td>39</td>
<td>53</td>
<td>50</td>
<td>32</td>
</tr>
<tr>
<td>B</td>
<td>48</td>
<td>36</td>
<td>82</td>
<td>32</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>G</td>
<td>56</td>
<td>45</td>
<td>30</td>
<td>79</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>R1</td>
<td>35</td>
<td>48</td>
<td>23</td>
<td>20</td>
<td>76</td>
<td>61</td>
</tr>
<tr>
<td>R2</td>
<td>32</td>
<td>42</td>
<td>14</td>
<td>29</td>
<td>65</td>
<td>79</td>
</tr>
</tbody>
</table>

Table 3.
Exact match rates for Obs. 2

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Y</th>
<th>B</th>
<th>G</th>
<th>R1</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>94</td>
<td>86</td>
<td>36</td>
<td>59</td>
<td>35</td>
<td>26</td>
</tr>
<tr>
<td>Y</td>
<td>83</td>
<td>95</td>
<td>26</td>
<td>42</td>
<td>35</td>
<td>21</td>
</tr>
<tr>
<td>B</td>
<td>36</td>
<td>27</td>
<td>91</td>
<td>24</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>G</td>
<td>68</td>
<td>44</td>
<td>35</td>
<td>77</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>R1</td>
<td>29</td>
<td>35</td>
<td>12</td>
<td>18</td>
<td>73</td>
<td>61</td>
</tr>
<tr>
<td>R2</td>
<td>30</td>
<td>26</td>
<td>9</td>
<td>18</td>
<td>53</td>
<td>70</td>
</tr>
</tbody>
</table>

Table 4.
Exact match rates for Obs. 3

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Y</th>
<th>B</th>
<th>G</th>
<th>R1</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>82</td>
<td>71</td>
<td>27</td>
<td>24</td>
<td>26</td>
<td>20</td>
</tr>
<tr>
<td>Y</td>
<td>74</td>
<td>92</td>
<td>18</td>
<td>21</td>
<td>33</td>
<td>27</td>
</tr>
<tr>
<td>B</td>
<td>30</td>
<td>23</td>
<td>76</td>
<td>33</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>G</td>
<td>42</td>
<td>24</td>
<td>21</td>
<td>74</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>R1</td>
<td>32</td>
<td>39</td>
<td>11</td>
<td>15</td>
<td>76</td>
<td>56</td>
</tr>
<tr>
<td>R2</td>
<td>23</td>
<td>20</td>
<td>12</td>
<td>11</td>
<td>56</td>
<td>73</td>
</tr>
</tbody>
</table>

First, we checked if the observers settings could be accounted for simply by chromatic adaptation of ‘von Kries type’ (Von Kries, 1905/1970) (i.e., cone response normalisation) taking place independently in the test and match fields. This model (referred to as the von Kries coefficient rule) predicts a linear relationship between the corresponding cone responses to the test and match papers. We evaluated the $i$th cone response, $\varphi_i(x)$, to a stimulus paper with the spectral reflectance function $x(\lambda)$ illuminated by a light with the spectral power distribution $I(\lambda)$ as

$$\varphi_i(x) = \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} x(\lambda) I(\lambda) s_i(\lambda) \, d\lambda,$$  \hspace{1cm} (1)
Table 5.
Exact match rates for Obs. 4

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Y</th>
<th>B</th>
<th>G</th>
<th>R1</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>94</td>
<td>71</td>
<td>47</td>
<td>39</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>Y</td>
<td>79</td>
<td>95</td>
<td>24</td>
<td>48</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td>B</td>
<td>45</td>
<td>27</td>
<td>88</td>
<td>33</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>G</td>
<td>56</td>
<td>52</td>
<td>32</td>
<td>80</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>R1</td>
<td>32</td>
<td>42</td>
<td>26</td>
<td>20</td>
<td>80</td>
<td>59</td>
</tr>
<tr>
<td>R2</td>
<td>33</td>
<td>33</td>
<td>20</td>
<td>18</td>
<td>73</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 6.
Exact match rates averaged across all the observers

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Y</th>
<th>B</th>
<th>G</th>
<th>R1</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>92</td>
<td>77</td>
<td>39</td>
<td>42</td>
<td>31</td>
<td>27</td>
</tr>
<tr>
<td>Y</td>
<td>80</td>
<td>93</td>
<td>27</td>
<td>41</td>
<td>39</td>
<td>28</td>
</tr>
<tr>
<td>B</td>
<td>40</td>
<td>28</td>
<td>84</td>
<td>31</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>G</td>
<td>56</td>
<td>41</td>
<td>30</td>
<td>78</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>R1</td>
<td>32</td>
<td>41</td>
<td>18</td>
<td>18</td>
<td>76</td>
<td>59</td>
</tr>
<tr>
<td>R2</td>
<td>30</td>
<td>30</td>
<td>14</td>
<td>19</td>
<td>62</td>
<td>75</td>
</tr>
</tbody>
</table>

Table 7.
Mean mismatch rate (MMMR) in hue steps

<table>
<thead>
<tr>
<th>Light</th>
<th>N</th>
<th>Y</th>
<th>B</th>
<th>G</th>
<th>R1</th>
<th>R2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMRM</td>
<td>1.09</td>
<td>1.17</td>
<td>2.01</td>
<td>2.14</td>
<td>1.81</td>
<td>2.12</td>
<td>1.86</td>
</tr>
</tbody>
</table>

Table 8.
Average match and mismatch rates. The acronym AMMR stands for achromatic mismatch rate; EMR for exact match rate; 1MMR, 2MMR and 3MMR for proportions of the matches for which mismatch was not more than one, two and three hue steps, respectively

<table>
<thead>
<tr>
<th>Light</th>
<th>AMMR</th>
<th>EMR</th>
<th>1MMR</th>
<th>2MMR</th>
<th>3MMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>3.17</td>
<td>40.5</td>
<td>70.5</td>
<td>84.5</td>
<td>92.0</td>
</tr>
<tr>
<td>Y</td>
<td>3.33</td>
<td>39.0</td>
<td>69.3</td>
<td>82.6</td>
<td>90.0</td>
</tr>
<tr>
<td>B</td>
<td>4.33</td>
<td>20.4</td>
<td>51.8</td>
<td>66.7</td>
<td>78.0</td>
</tr>
<tr>
<td>G</td>
<td>4.67</td>
<td>26.4</td>
<td>52.0</td>
<td>64.5</td>
<td>75.0</td>
</tr>
<tr>
<td>R1</td>
<td>3.54</td>
<td>28.6</td>
<td>56.8</td>
<td>72.1</td>
<td>80.8</td>
</tr>
<tr>
<td>R2</td>
<td>4.29</td>
<td>24.8</td>
<td>50.3</td>
<td>65.7</td>
<td>76.2</td>
</tr>
<tr>
<td>Total</td>
<td>3.89</td>
<td>29.9</td>
<td>57.8</td>
<td>70.5</td>
<td>78.8</td>
</tr>
</tbody>
</table>
where \( s_i(\lambda) \) is the spectral sensitivity of the \( i \)th cone-type \((i = 1, 2, 3)\), \( \lambda_{\text{min}} = 380 \text{ nm} \), \( \lambda_{\text{max}} = 780 \text{ nm} \). The cone spectral sensitivities \( s_i(\lambda) \) were modelled as a product of the photopigment template (Govardovskii et al., 2000) and the transmittance spectra of the lens and macular pigment (Wyszecki and Stiles, 1982). The photopigment templates with peak sensitivity at 425, 530 and 560 nm were chosen to model the short (S-), middle (M-) and long (L-) cones, respectively.
Figure 14. Each plate presents a scatter plot of the cone responses to the match (averaged over observers) vs. test papers for one illumination condition. The S-cone scatter plots are displayed in the left column, the M-cone in the middle column, and the L-cone in the right column. Only those illumination conditions where the test light was neutral are presented. The illumination of the match field is designated by the vertical axis label. The black lines represent simple linear regression. In the right bottom corner is the degrees of freedom adjusted $R$-square coefficient. The closer its value is to 1, the better fit is. This figure is published in color in the online version.

Figure 14 shows that the von Kries coefficient rule fails to predict matches made in our experiment even when the test light was neutral. Indeed, if the observers’ matches followed the von Kries coefficient rule prediction, the data points in Fig. 14 would be on a straight line in each plate. Therefore, the deviation from linearity in Fig. 14 shows the discrepancy between the data obtained and the von Kries coefficient rule based prediction. When the test lights are chromatic the matches are even less consistent with the von Kries coefficient rule. That the test and match cone responses are poorly correlated is hardly surprising since the cone ratios have not been found to be invariant for the stimulus papers and lights used in our experiment (Fig. 15). Indeed, in Fig. 15 the cone responses to the stimulus papers under all the five chromatic lights engaged in the experiment are plotted versus the
Figure 15. Scatter plots of the cone responses to the test papers calculated for the neutral, on the one hand, and one of the five chromatic lights (as designated in the vertical axis label), on the other. In each column the plot for the cone responses of one type are displayed as in Fig. 14. The black lines represent simple linear regression (with the degrees of freedom adjusted $R$-square coefficient in the right bottom corner). This figure is published in color in the online version.

cone responses to the same papers under the neutral light. As one can see, it is only switching from the neutral to yellow light (the spectral power distributions of which differ least of all) produces approximately linear changes in the cone responses of all the three types. The relationships between the corresponding cone responses when both lights are chromatic are far from linear.

As von Kries’ adaptation is incorporated in many colour appearance models, including those endorsed by the CIE (e.g., CIELAB, CIECAM02), our test is relevant to these models as well as to some proposed modifications of the von-Kries coefficient rule-based model (e.g., Funt and Ciurea, 2002).

We also tested a model put forward by Chichilnisky and Wandell (1995) which predicts a linear relationship between the decremental cone responses to the test and match papers. More specifically, they suggested that a match between a target
against one background and another target against another background occurs when for the cone response of each type the following equation holds:

$$g_{S_i} (S_{B_i} - S_{T_i}) = g_{S_i} (S_{B_2} - S_{T_2}),$$  \hspace{1cm} (2)$$

where $S_{B_i}$ is the S- (respectively, M- or L-) cone responses to the $i$th background; $S_{T_i}$ is the S- (respectively, M- or L-) cone responses to the $i$th target; and $g_{S_i}$ stands for the gain of the S- (respectively, M- or L-) cones determined by the $i$th background; $i = 1, 2$. Figure 16 shows that this prediction fails too. Although Chichilnisky and Wandell contrasted their model to von Kries’ coefficient rule, the graphs in Figs 14 and 16 look surprisingly alike.

Next, we tested a model predicting equal colour appearance for chromatic targets presented against different chromatic backgrounds when the differences between the logarithms of the corresponding cone responses to the target and background
Figure 17. Same as Fig. 14 except that the decremental logarithmic cone responses are used. The decremental logarithmic cone response to the paper under the neutral light is on the horizontal axis, that under the chromatic light being on the vertical axis. In each column the plot for the cone responses of one type are displayed as in Fig. 14. The black lines represent simple linear regression (with the degrees of freedom adjusted R-square coefficient in the right bottom corner). This figure is published in color in the online version.

are equal (Shepherd, 1997; Whittle, 2003). Figure 17 testifies against this model as well. To summarise, the standard approaches based on the various types of differencing and normalisation of the cone excitation values (e.g., Ebner, 2007; Hurlbert, 1998) prove to be unsuccessful in accounting for our results. It must be borne in mind, however, that the cone excitation space is not quite appropriate for representing the colour of objects under varying illumination: nor is any colorimetric space based on trichromatic colour matching functions (Wyszecki and Stiles, 1982). In fact, these spaces are supposed to represent the colour of the light reflected from objects rather than the colour of the objects themselves. As the same object reflects different light when illumination varies, the object representation in these spaces
alters with illumination. It is not always clear whether this alteration is due to a change of the object colour, or it is simply an artifact of the illumination change.

We believe that a new colour space purposely designed to represent object colour (Godau and Funt, 2011; Logvinenko, 2009b) is more suitable for analysing the results of these colour constancy experiments. The novelty of this space is in using a special set of rectangular spectral reflectance functions (Fig. 18) as a common frame reference for the representation of object colour under varying illumination. It has been proven that under fixed illumination for any spectral reflectance function there is exactly one rectangular spectral reflectance function to which it is metameric (see Note 4). We will refer to the latter as a rectangular metamer of the former. Therefore, the rectangular spectral reflectance functions in Fig. 18 can represent all the object colours under a fixed illumination. Moreover, as the set of rectangular metamers is independent of illumination it can be used for representing object colours for all illuminations. Admittedly, it may happen that the same spectral reflectance function may have different rectangular metamers under different illuminations. This means that the object colour as such may change with illumination (for more details see Logvinenko, 2009b).

Each rectangular spectral reflectance function can be specified by three numbers: purity (\(\alpha\)), spectral band (\(\delta\)) and central wavelength (\(\lambda\)) (see Note 5) (Fig. 18). It has been shown that Munsell chips of the same central wavelength have practically the same Munsell Hue (see Note 6) (Logvinenko, 2009b). In other words, central wavelength turns out to be a stimulus correlate of Munsell Hue. Spectral band relates to the amount of whiteness/blackness. Purity correlates with the perceptual strength of the chromatic quality determined by the spectral band and central wavelength.

As mentioned above, a particular spectral reflectance function can have different rectangular metamers under different illuminations. Following the terminology put forth in Logvinenko (2009b) we will refer to this phenomenon as the \textit{illuminant induced colour stimulus shift} (Logvinenko, 2009b). The illuminant induced colour stimulus shift results from the so-called mismatching of metamers that occurs when two objects metameric under one illumination cease being such under the other (Wyszecki and Stiles, 1982).
It seems plausible to expect the same rectangular spectral reflectance function to be assigned the same material colour under different illuminations (see Discussion). If also the least dissimilar match is based on the equality of material colours then we can make a prediction for our stimulus papers evaluating the colour stimulus shift produced by the illuminants used in our experiment. The prediction is rather simple: the least dissimilarity between differently illuminated papers is to be achieved by the pair with the same rectangular metamers. As purity and spectral band did not vary systematically over the stimulus sample, this suggestion amounts, at first approximation, to the prediction that in our experiment the least dissimilar match should be determined by the central wavelength.

In order to test this hypothesis, rectangular metamers for all the stimulus Munsell papers have been evaluated using the cone fundamentals based on the photopigment template of Govardovskii and collaborators (Govardovskii et al., 2000) as described by Logvinenko (2009b). Table 9 displays the central wavelengths of the Munsell papers for all the six illuminants used in the experiment. As expected, the central wavelengths of each paper were found to be somewhat different for different lights (see Note 7).

The central wavelength shift can be observed in Figs 19–25 where the horizontal and vertical axes are the central wavelength of the papers under the test and match lights, respectively. The twenty-two black crosses in each plate mark the stimulus

Table 9.
Central wavelengths (nm)

<table>
<thead>
<tr>
<th>Light</th>
<th>Paper</th>
<th>5RP</th>
<th>10P</th>
<th>5P</th>
<th>10PB</th>
<th>5PB</th>
<th>10B</th>
<th>5B</th>
<th>10BG</th>
<th>5BG</th>
<th>10G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>N</td>
<td>722</td>
<td>735</td>
<td>744</td>
<td>461</td>
<td>476</td>
<td>480</td>
<td>485</td>
<td>489</td>
<td>495</td>
<td>499</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>721</td>
<td>734</td>
<td>742</td>
<td>455</td>
<td>478</td>
<td>483</td>
<td>487</td>
<td>492</td>
<td>498</td>
<td>501</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>458</td>
<td>454</td>
<td>456</td>
<td>460</td>
<td>465</td>
<td>471</td>
<td>479</td>
<td>490</td>
<td>499</td>
<td>505</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>472</td>
<td>479</td>
<td>484</td>
<td>488</td>
<td>495</td>
<td>497</td>
<td>499</td>
<td>502</td>
<td>506</td>
<td>510</td>
<td></td>
</tr>
<tr>
<td>R1</td>
<td>632</td>
<td>730</td>
<td>741</td>
<td>754</td>
<td>773</td>
<td>492</td>
<td>496</td>
<td>503</td>
<td>513</td>
<td>515</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>640</td>
<td>732</td>
<td>740</td>
<td>754</td>
<td>472</td>
<td>480</td>
<td>476</td>
<td>488</td>
<td>496</td>
<td>498</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Paper</th>
<th>5G</th>
<th>10GY</th>
<th>5GY</th>
<th>10Y</th>
<th>5Y</th>
<th>10YR</th>
<th>5YR</th>
<th>10R</th>
<th>5R</th>
<th>10RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>509</td>
<td>545</td>
<td>562</td>
<td>571</td>
<td>577</td>
<td>583</td>
<td>593</td>
<td>605</td>
<td>639</td>
<td>715</td>
</tr>
<tr>
<td>12</td>
<td>510</td>
<td>549</td>
<td>565</td>
<td>573</td>
<td>579</td>
<td>585</td>
<td>594</td>
<td>604</td>
<td>629</td>
<td>714</td>
</tr>
<tr>
<td>13</td>
<td>510</td>
<td>519</td>
<td>529</td>
<td>532</td>
<td>535</td>
<td>539</td>
<td>541</td>
<td>555</td>
<td>739</td>
<td>451</td>
</tr>
<tr>
<td>14</td>
<td>519</td>
<td>536</td>
<td>544</td>
<td>551</td>
<td>550</td>
<td>553</td>
<td>558</td>
<td>569</td>
<td>709</td>
<td>719</td>
</tr>
<tr>
<td>15</td>
<td>522</td>
<td>599</td>
<td>612</td>
<td>616</td>
<td>618</td>
<td>619</td>
<td>622</td>
<td>624</td>
<td>627</td>
<td>634</td>
</tr>
<tr>
<td>16</td>
<td>503</td>
<td>583</td>
<td>599</td>
<td>603</td>
<td>608</td>
<td>610</td>
<td>614</td>
<td>617</td>
<td>623</td>
<td>629</td>
</tr>
</tbody>
</table>
Figure 19. Median least dissimilar matches made by four observers when the test and match fields were illuminated by the same light. The central wavelength for test and match papers are plotted on the horizontal and and vertical axes, respectively. (As the central wavelength is not defined for achromatic papers, we arbitrarily assigned the values 1.1 and 1.2 for the grey and black papers, respectively.) Different markers represent different observers. The black crosses indicate physically exact matches. The letters in the brackets in the axes labels designate the illuminant. This figure is published in color in the online version.

Munsell papers. Specifically, the abscissa of a cross is the central wavelength of a paper under the test illumination, and the ordinate is the central wavelength of the same paper under the match illumination. The deviation from the bisection indicates the magnitude of the central wavelength shift under given pair of lights. Note that due to rescaling, central wavelength varies within the interval [0; 1]. The end points of the interval correspond to the visible spectrum interval ends, with the short and long wavelength ends corresponding to 0 and 1, respectively (for more details see Logvinenko, 2009b).

The black diagonal in each plate indicates the prediction based on the central wavelength equality. Indeed, if this prediction is correct then the central wavelengths of the test and match papers must be close if not equal (see Note 8).

In Figs 19–25 the median matches of all the four observers for all the 36 illumination conditions are presented. Each marker stands for the median (see Note 9) of the 3 matches made by one observer for each paper. The predictions based on the physically exact match (i.e., when the match paper is the same as the test paper) are the black crosses. As may be seen, the observers’ matches drastically violate
Figure 20. Median least dissimilar matches made by four observers when the test field was illuminated by the neutral light. The match field illumination is marked by the letter in the brackets in the vertical axis labels. The central wavelength for test and match papers are plotted on the horizontal and and vertical axes, respectively. Different markers represent different observers. The black crosses indicate physically exact matches. This figure is published in color in the online version.

the central wavelength equality prediction. Moreover, the actual matches gravitate more to the black crosses rather than to the diagonal. Close inspection of the graphs shows that in many cases the observers chose as least dissimilar a paper physically identical to the test paper despite the fact that in the match field there were papers the central wavelength of which was much closer (sometimes practically equal) to the central wavelength of the test paper.

5. Discussion

Textbooks usually describe colour constancy as relative stability of the colour of an object despite changes of the object illumination. Such a definition depends on what exactly is meant by ‘relative stability’ and ‘object colour’. Is object colour subject to the same relative stability as that of the gravitational acceleration on Earth which is never observed under natural conditions but which can be observed under ideal conditions (i.e., in a vacuum)? Or is it rather similar to the ‘economic stability’ which cannot be achieved even in principle? In other words, does the adjective ‘relative’ in the definition of the colour constancy conceal the fact that colour constancy
is something that cannot exist in principle? We incline to the affirmative answer to this question for a number of reasons.

First of all, as many colour scientists emphasised before (e.g., Evans, 1948; Fairchild, 2005, p. 132; Hunt, 1977), the colour appearance of an object does change with illumination. Modern colour appearance models, such as CIECAM02, operate with more than three colour dimensions, attempting to take into account the effect of illumination on colour appearance (Fairchild, 2005). Driven by similar intentions, we have recently suggested distinguishing between material and lighting object colour dimensions (Tokunaga and Logvinenko, 2010a, b, c). When the illumination changes, the colour of an object changes mainly in the lighting colour dimensions. For instance, white surfaces are known to look white irrespective of illumination. However, this does not mean that their colour appearance does not change with illumination. Indeed, two adjacent white surfaces — one lit by daylight, another by a red light — would certainly appear different despite the fact that observers readily realise that it is the same white surface. It is to capture this duality of colour appearance of reflecting objects in multiple light scenes that the distinction between material and lighting colour dimensions has been put forward. We believe that white surfaces under different illuminations have the same material
colour but differ in lighting colour. If this is the case then one can say that for white surfaces colour constancy takes place in terms of material colour dimensions. In other words, colour constancy is impossible if it is understood literally, that is, as absolute stability of the colour appearance of an object under different illuminations. Nevertheless, it might be possible in the sense that a change in illumination affects only the lighting colour, the material colour remaining constant.

It appears that in our experiment colour constancy in terms of material colour dimensions took place for the two achromatic papers. Indeed, in 71% of the trials with the grey paper, observers decided that it was the same grey paper in the matching field that looked least dissimilar to this paper in the test field. It does not mean that the two grey papers had the same colour appearance when the illumination of the test and match fields was different. There was a clear apparent difference between these two grey papers produced by the difference in illumination, that is, the difference in terms of the lighting colour dimensions. Because of this lighting difference there was no paper at all in the matching field which could bring exact match to the grey paper in the test field. Therefore, exact asymmetric colour match was impossible in principle because of the existence of the lighting colour dimensions of object colour. Still, the grey paper in the matching field was consistently

Figure 22. Median least dissimilar matches made by four observers when the test field was illuminated by the blue light. The match field illumination is marked by the letter in the brackets in the vertical axis labels. The central wavelength for test and match papers are plotted on the horizontal and and vertical axes, respectively. Different markers represent different observers. The black crosses indicate physically exact matches. This figure is published in color in the online version.
judged as least dissimilar from that paper in the test field as compared to all the other papers in the matching field. If the assumption that the minimal dissimilarity between differently illuminated papers is achieved when the papers have the same material colour (see Note 10) holds true, then the 71% exact match rate for the grey paper indicates that our observers exhibited quite high material colour constancy with respect to illumination for the grey paper.

Thus, the visual system was found to be able to disentangle the reflected light reaching the eye (varied from one illumination condition to the other) so as to exhibit the material colour constancy for the grey paper. How does the visual system extract the material and lighting colour dimensions from the reflected light? The classical approach, descended from Helmholtz (Helmholtz, 1867), is based on the idea that the visual system estimates somehow the illuminant so as to use this estimate to derive the ‘true’ colour of the object. It must be said that it is not quite clear what Helmholtz understood by the ‘true’ colour of an object, and what kind of illuminant estimation he had in mind. Yet, the illuminant estimation idea is at the heart of the computational approach to colour constancy (for a review, see, e.g., Ebner, 2007). In the computational context the ‘true’ colour of an object is usually understood as the colour of the light reflected from the object when it is lit by the
Figure 24. Median least dissimilar matches made by four observers when the test field was illuminated by the red #1 light. The match field illumination is marked by the letter in the brackets in the vertical axis labels. The central wavelength for test and match papers are plotted on the horizontal and and vertical axes, respectively. Different markers represent different observers. The black crosses indicate physically exact matches. This figure is published in color in the online version.

‘neutral’ light. This is despite the fact that it is well known that for various reasons object colour cannot be reduced to light colour, not to mention the obvious arbitrariness of the choice of ‘neutral’ light. The goal of illuminant estimation is usually reduced to estimating the cone responses to the incident light (see Note 11). To be more exact, within the computational approach to colour constancy the major goal is understood as follows: given the cone responses to the light reflected from an object, to estimate the cone responses to the incident light, and then to derive the cone responses to the light which would be reflected by the object should it be lit by the ‘neutral’ light. In this context the major problem is to assess the illuminant cone responses. A number of computational algorithms have been proposed for illuminant estimation (for review see, e.g., Barnard et al., 2002; Ebner, 2007; Gijsenij et al., 2010; Hordley, 2006; Hurlbert 1998; Maloney, 1999). Still, there is no consensus on which of these, if any, makes the right predictions about human perception. There has been little testing of this question because there are not good quantitative results of human perception on complex scenes.

It should be borne in mind that there is an infinite number of various spectral distribution functions with the same triplet of the cone responses, which produce different reflected light with different cone responses for the same reflecting object.
Figure 25. Median least dissimilar matches made by four observers when the test field was illuminated by the red #2 light. The match field illumination is marked by the letter in the brackets in the vertical axis labels. The central wavelength for test and match papers are plotted on the horizontal and vertical axes, respectively. Different markers represent different observers. The black crosses indicate physically exact matches. This figure is published in color in the online version.

Thus, the illuminant estimation task as it stands in the computational approach is an ill-posed problem, that is, there is no unique solution for it. This uncertainty is usually either ignored, or implicitly thought to be not important. The general belief that an ‘approximate’ illuminant estimate will suffice, can be expressed by paraphrasing Shevell and Kingdom, (2008, p. 150): estimation is not perfect but neither is colour constancy.

Interestingly, if we take for granted that the material colour constancy achieved by our observers for the grey paper results from illuminant estimation, one has to conclude that this estimation must have been rather accurate because the exact match rate for the grey paper was quite high (71%). The following questions immediately arise then. First, why this illuminant estimate was not used to achieve constant material colour perception for other papers? Second, why the material colour inconstancy for the chromatic papers observed in our experiment was so different for different papers?

One possible reason might be the way the visual system might use the hypothetical illuminant estimation. There is a strong belief that it is used to normalise the cone responses to the light reflected from an object in order to estimate its ‘true’ colour (see, e.g., Brainard, 2009b; Shevell and Kingdom, 2008). That such
estimation based on the simple (von Kries) cone response normalisation is also ‘ap-
proximate’ is well known (see Note 12). However, we found very little support for
the cone-response-normalisation account of our data.

The belief in the von Kries normalisation has been supported by the claim of
the cone ratio invariance with respect to illumination (Dannemiller, 1989; Smith-
son, 2005; Zaidi, 2001). As only particular samples of reflectance spectra have been
tested by far, we decided not only to ascertain cone ratio variance for Munsell pa-
pers, but also to evaluate the theoretical limits of the cone response variance for
all the reflectance spectra (see Appendix A.1). As one can see in Fig. 26, the cone
response variability is rather large even for the Planckian radiators with the colour
temperature 2000 K and 100,000 K, not to mention the chromatic illuminants used
in the present study; though this is in line with the previous finding that “cone ex-
citations are themselves not invariant under changes in natural illuminant” (Foster
and Nascimento, 1994, p. 116). In fact, Foster and Nascimento (1994) claimed that
it was spatial ratios of cone excitations, not cone excitations themselves, that were
relatively invariant with respect to illuminant changes. Yet we found that the cone
ratios were as inconstant as the cone responses for Munsell papers (Fig. 27).

Thus, we conclude that being inconstant, neither cone responses, nor cone re-
sponse ratios can, generally, secure constant colour perception.

Moreover, we believe that the very understanding of the colour constancy phe-
nomenon as recovering the true (intrinsic) colour of an object irrespective of its
illumination is at least misleading because there is no true colour for most real ob-
jects. Objects usually do change their colour with illumination. To show this we
have to specify what we mean by object colour.

When an illuminant is fixed, object colour can be defined (from the psychophys-
ical point of view) as a class of metameric spectral reflectance functions, the object
colour palette being defined as a set of classes of metameric spectral reflectance
functions. Each class comprises all the spectral reflectance functions producing the
same triplet of the cone responses to the reflected light. Two surfaces \( x_1(\lambda) \) and
\( x_2(\lambda) \) illuminated by the same light are said to be metameric (and have the same
object colour) if, and only if, in equation (1) \( \varphi_i(x_1) = \varphi_i(x_2) \) for each \( i \).

When \( \varphi_1, \varphi_2, \varphi_3 \) are treated as the Cartesian coordinates, and \( x(\lambda) \) runs through
all the spectral reflectance functions, the cone responses make a solid (referred to
as an object colour solid) in the 3D space. Therefore, in this space the object colour
palette for a fixed illuminant is geometrically represented by the object colour solid.
The object colour solids for different illuminants have different shape.

The observers’ ability to make an ‘approximate’ match between differently il-
uminated objects would be impossible unless observers were able to put in cor-
respondence object colours in different object colour palettes. Let us call this hy-
pothetical correspondence the across-illuminant colour map. In other words, the
across-illuminant colour map is the map between the object colour palettes pro-
duced under two different illuminations. Being built in somewhere in the brain,
such a map underlies asymmetric colour matching across two illuminations.
Figure 26. Cone responses to 1600 Munsell papers illuminated by the Plankian radiator with the colour temperature 2000 K versus 100,000 K. Cone responses have been normalised by the maximum response. The smooth contour confines the area which would be filled up if the cone responses were evaluated for all the possible reflecting objects. This figure is published in color in the online version.

One can conjecture that two object colour stimuli put in correspondence via the across-illuminant colour map have the same material colour differing only in lighting dimensions. If this is the case, and if the across-illuminant colour map underlies
the least dissimilar matching, then the least dissimilar matching must be, firstly, symmetric, and secondly, in a sense, transitive. While it seems natural to expect the least dissimilar matching to be symmetric, the transitivity is a rather strong prediction since dissimilarity, as well as similarity, is, generally, thought of as an intransitive binary relation (e.g., Tversky, 1977).
Table 10.
Mean transitivity rates

<table>
<thead>
<tr>
<th></th>
<th>Obs. 1</th>
<th>Obs. 2</th>
<th>Obs. 3</th>
<th>Obs. 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.60</td>
<td>0.55</td>
<td>0.54</td>
<td>0.56</td>
<td>0.56</td>
</tr>
</tbody>
</table>

By transitivity of the least dissimilar matching in the present context we understand the following. If an object colour stimulus $x_1$ illuminated by a light $p_1$ is judged as least dissimilar to an object colour stimulus $x_2$ illuminated by a light $p_2$ (that is, $x_1$ and $x_2$ have the same material colour), written $(x_1, p_1) \sim (x_2, p_2)$; and if, in turn, the object colour stimulus $x_2$ illuminated by the light $p_2$ is judged as least dissimilar to an object colour stimulus $x_3$ illuminated by a light $p_3$ (that is, $x_2$ and $x_3$ also have the same material colour), written $(x_2, p_2) \sim (x_3, p_3)$; then the object colour stimulus $x_1$ illuminated by a light $p_1$ must be judged as least dissimilar to the object colour stimulus $x_3$ illuminated by the light $p_3$ (because $x_1$ and $x_3$ have the same material colour) written $(x_1, p_1) \sim (x_3, p_3)$.

To what extent are our experimental results consistent with the transitivity hypothesis? Given a triad of lights $p_1$, $p_2$ and $p_3$, let us say that a triplet $(x_1, p_1), (x_2; p_2)$ and $(x_3; p_3)$ is favourable to transitivity if $(x_1, p_1) \sim (x_2, p_2)$, and $(x_2, p_2) \sim (x_3, p_3)$, and $(x_1, p_1) \sim (x_3, p_3)$; and contradicting transitivity if $(x_1, p_1) \sim (x_2, p_2)$, and $(x_2, p_2) \sim (x_3, p_3)$, but $(x_1, p_1) \not\sim (x_3, p_3)$. Let us call transitivity rate the number of triplets favourable to transitivity over the combined number of triplets favourable and contradicting transitivity. We evaluated the transitivity rate for each of the 216 possible triads of the lights used in the experiment, and every observer. The mean transitivity rates are presented in Table 10. As may be seen, on average the results are quite favourable to the transitivity hypothesis. However, the significance of this can be established only by using a proper statistical procedure. Using the statistical test designed on purpose to test transitivity (Logvinenko et al., 2004), we found that the null hypothesis of intransitivity could be rejected with high significance for each observer (see Appendix A.2). Therefore, the occurrence of the triplets contradicting transitivity can be accounted for by independent random fluctuations of observers’ responses. Applying a similar test we also established the symmetry of the least dissimilar matching (see Appendix A.2).

The transitivity of least dissimilar matching shows that the deviations from exact match are systematic — that they can hardly be interpreted as a result of unbiassed random errors due to inaccurate performance. It justifies the notion of an across-illuminant colour map which will allow us to lead colour research out of the deadlock which it has reached due to the colour constancy metaphor. The notion of colour constancy predetermines and impoverishes the very conceptual framework of colour research. Indeed, if one believes in colour constancy, then only two possible outcomes can be expected: either colour constancy, or colour inconstancy. The only subject left to experimental research in this case is the degree of colour incon-
stancy. However, more than half a century of extensive studies of colour constancy has brought no consensus. Some authors claim that colour constancy is rather high (Arend and Reeves, 1986; Arend et al., 1991; Bauml, 1999; Brainard, 1998), some that it is very poor (Evans, 1948; Helson and Judd, 1936), and some that it depends on illumination, circumstances of viewing, and other experimental conditions (Foster, 2011; Maloney, 2003).

An overriding notion, the notion of across-illuminant colour map opens a novel avenue for studying object colour perception. It explicitly suggests exploring the transformations which object colours undergo due to illumination changes rather than the degree of deviation from colour constancy which does not actually exist. Furthermore, colour constancy simply cannot exist. The fundamental fact is that there is no across-illuminant colour map which can secure exact (asymmetric colour) matches for all objects. Indeed, from the formal point of view the across-illuminant colour map can be considered as a map between different object colour solids. Generally, there is an infinite number of one-to-one maps between two arbitrary object colour solids. Which of them can be considered as likely candidates for the across-illuminant colour map? A natural requirement would be maximising the exact match rate. For instance, the surface reflecting no light at all (referred to as the absolute black) is represented by the zero triplets of the cone responses (referred to as the south pole of the object colour solid) for any illuminant. Moreover, if the illuminant is everywhere positive then it is only the absolute black that maps to the object colour solid south pole because there are no metamers for the absolute black for such illuminants (Logvinenko, 2009b). Therefore, if the across-illuminant colour map puts the south poles of the object colour solids in correspondence this will guarantee an exact match for the absolute black.

By the same line of reasoning we come to the conclusion that the across-illuminant colour map has to put in correspondence the north poles of the object colour solids. Indeed, it has been proved that for everywhere positive illuminants, the perfect reflector (i.e., the ideal white) is the only inverse image of (i.e., surface mapping to) the north pole of the object colour solid (Logvinenko, 2009b). Therefore, the across-illuminant colour map in this case can be considered as follows: the north pole \#1 \rightarrow its inverse image (the perfect reflector) \rightarrow the north pole \#2 (here the arrow ‘\rightarrow’ stands for ‘maps to’).

Generally, if the inverse images of two points in the two object colour solids coincide, then these object colour solid points should be put in correspondence by the across-illuminant colour map. Indeed, any other correspondence can bring about only lower exact match rates for the spectral reflectance functions in these inverse images. As shown elsewhere, the inverse image is a singleton not only for the north and south poles but for every boundary point of the object colour solid, provided that the illuminant is everywhere positive (Logvinenko, 2009b). The spectral reflectance functions mapping to the object colour solid boundary are called optimal object colour stimuli. It turns out that for everywhere positive illuminants, the optimal object colour stimuli cannot be metameric. Hence, in order to secure an
exact match for the optimal object colour stimuli, the across-illuminant colour map should put in correspondence those object colour-solid boundary points that have the same inverse image.

Yet, no across-illuminant colour map can secure an exact match for those spectral reflectance functions that map inside the object colour solids. The inverse images of two interior points in two object colour solids never coincide. They either partly intersect or do not intersect at all. The across-illuminant colour map can bring about exact matches only for the spectral reflectance functions inside the intersection of the inverse images. Those spectral reflectance functions which are outside the inverse images’ intersection undergo mismatching of metamers. Therefore, such spectral reflectance functions will be perceived to be of a different colour compared to that assigned to the spectral reflectance functions inside the intersection of the inverse images. Therefore, exact matches will be impossible for all of them. It follows that even material colour constancy is impossible to achieve for most of the spectral reflectance functions because of mismatching of metamers. The fact that the mismatching of metamers imposes certain limits on colour constancy has been acknowledged in the colour literature before (e.g., Fairchild, 2005, p. 132).

In this context the across-illuminant colour map becomes a key issue since it determines how inconstant the colour of a particular object will be under illumination alteration. It turns out that the way the visual system maps the object colour solids’ boundaries onto each other predetermines the entire across-illuminant colour map. Indeed, a remarkable property of human trichromatic colour vision is that each colour can be decomposed into a chromatic quality (hue (see Note 13)) and the subjective strength of this chromatic quality (referred to as hue intensity (see Note 14)). The points in the object colour solid lying on the same radius (from the object colour solid centre) prove to have approximately the same hue differing only in hue intensity (Logvinenko, 2009b). In other words, the contours of equal hue have a form of slightly curved radii. Let us call a set of object colour-solid points differing only in hue intensity a hue fiber. Then, the object colour solid can be thought of as a fan-like bundle of hue fibers that are approximately linear radii.

Therefore, if the two boundary points of the two different object colour solids are put in correspondence by the across-illuminant colour map, then the corresponding hue fibers should be put in correspondence by the across-illuminant colour map as well. Indeed, any other correspondence will only increase the dissimilarity because of the difference in material hue. It follows that the across-illuminant colour map is actually predetermined by its restriction to the object colour solid boundary. Let us call this restriction the boundary colour map.

The boundary colour map can be theoretically evaluated for any two illuminants. The hue fibers can be evaluated by experiment. Given the hue fibers, one can compute the across-illuminant colour map. This facilitates the prediction of asymmetric colour matches providing that it is mediated by the across-illuminant colour map based on the boundary colour map. The prediction based on the equality of the test and match central wavelengths is a prediction of this sort made under the as-
sumption of hue-fiber linearity (i.e., assuming that they are linear radii of the object colour solid).

Surprisingly, it was established that the observer matches in our experiment could not be accounted for by the central wavelength equality. Furthermore, we found that our results do not just deviate from the predictions of this hypothesis: the deviations systematically gravitated towards a physically exact match. That is, our observers were found to be much more successful in establishing exact matches than follows from the central-wavelength-equality hypothesis.

The central-wavelength-equality hypothesis is based on an assumption that the rectangular spectral reflectance functions are assigned the same material colour irrespective of illumination. In other words, the central-wavelength-equality hypothesis rests upon an assumption of material colour constancy for the rectangular spectral reflectance functions. That our observers exhibited, paradoxically, more material colour constancy (i.e., the higher exact match rate) than predicted by the central-wavelength-equality hypothesis means that their colour perception of the rectangular spectral reflectance functions would be inconstant. In other words, performing better than predicted for the set of Munsell papers selected for the experiment, our observers would, in theory, perform worse than predicted for the optimal object colour stimuli. Indeed, as the central-wavelength-equality prediction is based on the boundary colour map, it implies an exact match for all the optimal object colour stimuli. The failure of the central-wavelength-equality prediction entails inexact match for the optimal object colour stimuli.

Alternatively, the failure of the central-wavelength-equality prediction can indicate an essential curvilinearity of the hue fibers. The iso-hue contours as measured in the CIE chromaticity diagram for colours of lights are rather curved (Burns et al., 1984; Pridmore, 2007). Unfortunately, no systematic data have been collected for object colours, that is, we do not know the form of the analogous contours in the colour solid (i.e., the hue fibers). Hence, the clear next step is to establish experimentally the set of the hue fibers for various illuminations, and then to use these for prediction of the least dissimilar asymmetric colour matching.

In conclusion, note that although material colour constancy is not achievable for all the reflecting objects, it can be achieved for a particular sample of reflecting objects. Moreover, any across-illuminant colour map brings about perfect material colour constancy for some reflecting objects. Because some objects and lights are likely to be ecologically more valid than others, the across-illuminant colour map might have been developed so that it (i) secures full material colour constancy for some objects; and (ii) minimises the deviation from it for other objects and lights. For instance, our results indicate that the visual system seems to sacrifice the accuracy of asymmetric colour matching for ideal optimal object colour stimuli in favour of real surfaces such as Munsell papers.

Also, it should be taken into consideration that some reflecting objects are more likely to appear in the environment of an animal than the others. The across-illuminant colour map might have evolved so that it secures high material colour
constancy not only for more ecologically valid but also for more probable (to appear in the environment) reflecting objects. This opens an avenue for applying the Bayesian approach (Brainard, 2009a; Brainard et al., 2006) to model the across-illuminant colour map and, thus, material colour constancy.

6. Conclusion

When the illumination is constant the colour of an object can be described in terms of three dimensions. However, one needs more dimensions to describe the colour appearance of objects in multiple light scenes. Respectively, we distinguish between lighting and material dimensions of object colour. A change in illumination causes alterations of the colour appearance of an object of two kinds. The major alteration is experienced in terms of the lighting dimensions of object colour. It is because of this unavoidable lighting difference that perfect asymmetric matching across illumination change is, strictly speaking, impossible. At first glance, the colour of an object is rather robust as far as the material colour dimensions are concerned. Still, our theoretical analysis shows that because of metamer mismatching the illuminant change induces systematic alterations of object colour in terms of the material dimensions as well. Moreover, these alterations are different for different reflecting objects and different illuminants. Our results corroborate these theoretical predictions. Specifically, the deviations from material colour constancy observed in our experiment are found to be different (i) for different reflecting objects under fixed illumination conditions; and (ii) for different illumination conditions for the same reflecting object.

Using the least dissimilar matching technique, we show that despite the deviations from exact matches, asymmetric colour matching is systematic. In particular, it is shown to be symmetric and transitive. This allows us to introduce a notion of an across-illuminant colour map that relates material colours under various illuminations. We argue that the across-illuminant colour map is a fundamental characteristic of human colour vision that has yet to be studied.

Acknowledgements

This work was supported by a research grant EP/C010353/1 from EPSRC to AL. We wish to thank B. Funt for reading the manuscript, and valuable comments.

Notes

1. By the palette of object colours we mean here the nomenclature of all discernible colours one can experience while observing reflecting objects.

2. It should be mentioned that some authors have used a hidden chromatic light source in order to manipulate object colour (Brainard et al., 1997; De’Almeida et al., 2004). Their rationale was as follows. If observers mistake a variation in the illumination colour for a variation in object colour (i.e., if the chromatic
version of the Gelb effect occurs) then one can reduce the problem of manipulating the object colour to that of simply manipulating the illumination colour. The problem with this approach is that one cannot be sure that the Gelb effect always takes place in full. In other words, one cannot exclude the possibility that in different trials the chromatic Gelb effect will be of different strength. In any case, the idea of measuring colour constancy with a technique based on a phenomenon of ultimate colour inconstancy (i.e., the chromatic Gelb effect) seems rather controversial. So, choosing between two evils — a hidden chromatic light versus a discrete stimulus set of limited size — we decided in favour of the latter.

3. By hue step we mean here Munsell Hue step. In other words, a deviation by one hue step means that the observer picked out the Munsell paper adjacent to the test paper in the Munsell hue circle.

4. Two spectral reflectance functions are called metameric (under some illumination) if they reflect metameric lights.

5. For constant spectral reflectance functions, which take the same value across the entire visible spectrum, central wavelength is not defined. Objects with such reflectance functions look achromatic.

6. When the illumination is neutral the central wavelength values prove to be very close to the dominant wavelength values of the reflected light.

7. As pointed out above, the difference between central wavelengths should be estimated in terms of circular metrics. For example the central wavelengths 721 and 458 nm are close to the opposite ends of the visible spectrum. So the distance between them in the visible spectrum circle is much smaller than in the visible spectrum interval.

8. Admittedly, alongside central wavelength purity and spectral band also contribute to dissimilarity judgements. Yet, there is every indication that for the particular sample of Munsell papers used in the experiment it is the difference in central wavelength that is crucial for dissimilarity judgements.

9. We used a circular median as defined in circular statistics (Fisher, 1996) because the circular metric is more appropriate for describing the proximity between the chromatic papers used in the experiment. Indeed, if the central wavelengths of the matches are, say, 0.9, 0.7 and 0.1, then the ordinary median is 0.7 (yellow) whereas the circular median is 0.9 (purple). As the paper with the central wavelength of 0.1 also looks purplish, the choice in favour of circular median looks justified. When one of three matches was achromatic the rest of two being chromatic, it was discarded, that is, the median was evaluated only for the two chromatic matches. Likewise, when there were two achromatic matches and one chromatic, the median was evaluated only for the two achromatic matches.
10. Intuitively, it seems plausible that an additional (material colour) difference can only increase the dissimilarity between the papers produced by the illumination difference.

11. Though, there are computational models which aim to derive the spectral power distribution of the illuminant with greater precision (e.g., D’Zmura and Iverson, 1993).

12. For example, see a colour inconstancy index introduced by Hunt (1998, pp. 128–129).

13. Note that here we make no distinction between achromatic and chromatic hues. In other words, although it is against the definition of hue endorsed by the CIE, in the present context black and white differ in hue. See more on this issue in Logvinenko and Beattie (2011).

14. The corresponding dimension for light colour is saturation.

References


**Appendix A.1 Theoretical Limits for the Cone Response Variation Induced by Illuminant Alteration**

Consider the $i$th cone responses, $\varphi_{i1}(x)$ and $\varphi_{i2}(x)$, to a spectral reflectance function $x(\lambda)$ under lights $I_1(\lambda)$ and $I_2(\lambda)$:

$$\varphi_{ij}(x) = \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} x(\lambda) I_j(\lambda) s_i(\lambda) \, d\lambda, \quad j = 1, 2. \quad (A.1)$$

Formally, $\varphi_{ij}(x)$ in equation (A.1) can be interpreted as the response of the cone with the spectral sensitivity $\phi_j(\lambda) = I_j(\lambda) s_i(\lambda)$ to the spectral reflectance function $x(\lambda)$ under equal energy light. When $x(\lambda)$ runs over all the spectral reflectance functions the pairs $(\varphi_{i1}, \varphi_{i2})$ fill in an area in a plane which has the shape of the 2D object colour solid determined by the $\phi_1(\lambda)$ and $\phi_2(\lambda)$. At the same time this area shows how much the $i$th cone responses, $\varphi_{i1}(x)$ and $\varphi_{i2}(x)$, to the same spectral reflectance functions $x(\lambda)$ under lights $I_1(\lambda)$ and $I_2(\lambda)$ covary with each other. Therefore, evaluating the theoretical limits for the cone response variation induced by an illuminant alteration can be reduced to evaluating the boundary contour of the 2D object colour solid determined by the $\phi_1(\lambda)$ and $\phi_2(\lambda)$. The smooth contours in Fig. 26 are the object colour boundaries evaluated for the Planckian radiators with the colour temperature 2000 K and 100,000 K as described by Logvinenko (2009b).

**Appendix A.2 Testing Statistically Transitivity and Symmetry of Least Dissimilar Matching**

Denote $(x, p)$ a paper $x$ illuminated by a light $p$. We will use denotation $\sim$ for the least dissimilar matching. That is, $(x_1, p_1) \sim (x_2, p_2)$ stands for that the paper $x_1$ illuminated by the light $p_1$ is judged as least dissimilar to the paper $x_2$ illuminated by the light $p_2$. Transitivity of $\sim$ means that

$$(x_1, p_1) \sim (x_2, p_2) \quad \& \quad (x_2, p_2) \sim (x_3, p_3) \Rightarrow (x_1, p_1) \sim (x_3, p_3). \quad (A.2)$$
When a triplet of pairs \((x_1, p_1), (x_2, p_2)\) and \((x_3, p_3)\) is tested, the eight possible outcomes can take place (Table A.1). Note that only one outcome, namely, \(R_2\) contradicts the transitivity statement (A.2). Our objective is to ascertain whether the proportion of the outcomes \(R_2\) observed in the experiment is compatible with (A.2) (i.e., it can be accounted for by random fluctuations of observer’s responses). It must be noted that statistical hypothesis testing is mainly performed for quantitative statements. It is, generally, not clear how to test statistically logical statements such as (A.2). Here we adopt an approach developed by Logvinenko et al. (2004). The main idea is that in the case of stochastic uncertainty of observer’s responses (e.g., when they are contaminated by binary noise) if transitivity holds true the occurrence of either premise in (A.2) should increase the likelihood of occurrence of the inference. Denoting by \(P\{A|B\}\) the probability of \(A\) under the condition \(B\), this can be written as

\[
P\{((x_1, p_1) \sim (x_3, p_3))|((x_1, p_1) \sim (x_2, p_2) & (x_2, p_2) \sim (x_3, p_3))\} > \max(P\{((x_1, p_1) \sim (x_3, p_3))|((x_1, p_1) \sim (x_2, p_2))\},
\]

\[
P\{((x_1, p_1) \sim (x_3, p_3))|((x_2, p_2) \sim (x_3, p_3))\}.
\]  

(A.3)

Designate \(A\) the event which occurs when the paper \(x_2\) illuminated by the light \(p_2\) is judged as least dissimilar to the paper \(x_3\) illuminated by the light \(p_3\) on condition that the paper \(x_1\) illuminated by the light \(p_1\) was judged as least dissimilar to the paper \(x_2\) illuminated by the light \(p_2\); and \(B\) the event which occurs when the paper \(x_1\) illuminated by the light \(p_1\) is judged as least dissimilar to the paper \(x_3\) illuminated by the light \(p_3\) on the same condition. Likewise, designate \(A'\) as the event which occurs when the paper \(x_1\) illuminated by the light \(p_1\) is judged as least dissimilar to the paper \(x_2\) illuminated by the light \(p_2\) on condition that the paper \(x_2\) illuminated by the light \(p_2\) was judged as least dissimilar to the paper \(x_3\) illuminated by the light \(p_3\); and \(B'\) as the event which occurs when the paper \(x_1\) illuminated by the light \(p_1\) is judged as least dissimilar to the paper \(x_3\) illuminated by the light \(p_3\) on the same condition. Then, inequality (A.3) can be decomposed into the following two:

\[
P\{AB\} > P\{A\} P\{B\},
\]  

(A.4)

\[
P\{A'B'\} > P\{A'\} P\{B'\}.
\]  

(A.5)
As one can see, the transitivity hypothesis (A.2) amounts to the probabilistic interdependence of the events \( A \) and \( B \), on the one hand, and \( A' \) and \( B' \), on the other hand (i.e., the probabilistic interdependence of the premises and inference in (A.2)). In the present context the probabilistic independence of either \( A \) and \( B \) or \( A' \) and \( B' \) implies no transitivity. Thus, we tested the two null hypotheses

\[
P\{AB\} = P\{A\}P\{B\}, \quad \text{(A.6)}
\]

\[
P\{A'B'\} = P\{A'\}P\{B'\} \quad \text{(A.7)}
\]

against the alternatives (A.4) and (A.5), respectively. In other words, we tested that in statement (A.2) the inference is probabilistically independent from both the premises. We accept the transitivity hypothesis (A.2) if both the null hypotheses (A.6) and (A.7) are rejected in favour of the alternatives (A.4) and (A.5), respectively.

Table A.2 presents a 2 × 2 contingency table associated with the events \( A \) and \( B \). In this table \( n_{11}, n_{12}, n_{21} \) and \( n_{22} \) are the number of times when the events \( AB, A\overline{B}, \overline{A}B \) and \( \overline{A}\overline{B} \) occurred, respectively. Note that in terms of Table A.1 \( AB = R_1, A\overline{B} = R_2, \overline{A}B = R_3 \) and \( \overline{A}\overline{B} = R_4 \). The totals are as usual: \( n_{1+} = n_{11} + n_{21} \); \( n_{2+} = n_{12} + n_{22} \); \( n_{1+} = n_{11} + n_{12} \); \( n_{2+} = n_{21} + n_{22} \); and \( n_{++} = n_{1+} + n_{2+} = n_{11} + n_{12} + n_{21} + n_{22} \). A similar 2 × 2 contingency table is associated with the events \( A' \) and \( B' \) (Table A.3) in which \( A'B' = R_1, A'\overline{B'} = R_2, \overline{A}'B' = R_5 \) and \( \overline{A}'\overline{B'} = R_6 \). Therefore, testing transitivity can be reduced to testing independence in a 2 × 2 contingency table (e.g., Kendall and Stuart, 1979). We chose to use Fisher’s exact test to examine independence in Tables A.2 and A.3 (Kendall and Stuart, 1979).

**Table A.2.** Contingency table associated with events \( A \) and \( B \)

<table>
<thead>
<tr>
<th></th>
<th>( B )</th>
<th>( \overline{B} )</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>( n_{11} )</td>
<td>( n_{12} )</td>
<td>( n_{1+} )</td>
</tr>
<tr>
<td>( \overline{A} )</td>
<td>( n_{21} )</td>
<td>( n_{22} )</td>
<td>( n_{2+} )</td>
</tr>
<tr>
<td>Totals</td>
<td>( n_{+1} )</td>
<td>( n_{+2} )</td>
<td>( n_{++} )</td>
</tr>
</tbody>
</table>

**Table A.3.** Contingency table associated with events \( A' \) and \( B' \)

<table>
<thead>
<tr>
<th></th>
<th>( B' )</th>
<th>( \overline{B'} )</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A' )</td>
<td>( n_{11} )</td>
<td>( n_{12} )</td>
<td>( n_{1+} )</td>
</tr>
<tr>
<td>( \overline{A}' )</td>
<td>( n_{21} )</td>
<td>( n_{22} )</td>
<td>( n_{2+} )</td>
</tr>
<tr>
<td>Totals</td>
<td>( n_{+1} )</td>
<td>( n_{+2} )</td>
<td>( n_{++} )</td>
</tr>
</tbody>
</table>
The contingency Tables A.2 and A.3 were evaluated for all the $6^3$ possible triplets of lights $(p_1, p_2, p_3)$. Given a particular triplet of lights $(p_1, p_2, p_3)$, 22 stimulus papers make $22^3$ triplets of pairs $(x_1, p_1), (x_2, p_2)$ and $(x_3, p_3)$. Then the number of each of the outcomes $R_1, \ldots, R_8$ for these $22^3$ triplets was evaluated separately for every observer. As each paper was tested in experiment 3 times for each illumination condition, the whole population of triplets of pairs for which the number of outcomes was calculated was $3 \times 22^3$.

For each observer 216 contingency Tables A.2 and A.3 were generated. For all observers a highly significant ($p < 0.0001$) contingency was found in each table, that is, the events $A$ and $B$ (as well as the events $A'$ and $B'$) were found to be not independent. To be more exact, for all observers the null (no transitivity) hypotheses (A.6) and (A.7) were rejected in favour of the alternatives (A.4) and (A.5) respectively for each triplet of lights $(p_1, p_2, p_3)$.

A similar approach was applied to test symmetry of least dissimilar matching. Specifically, we would like to test whether the following statement holds true for the least dissimilar matching $\sim$:

$$(x_1, p_1) \sim (x_2, p_2) \Rightarrow (x_2, p_2) \sim (x_1, p_1). \quad (A.8)$$

The same line of reasoning leads to the following probabilistic statement similar to equation (A.3):

$$P\{(x_2, p_2) \sim (x_1, p_1)|(x_1, p_1) \sim (x_2, p_2)\} > P\{(x_2, p_2) \sim (x_1, p_1)\}. \quad (A.9)$$

Likewise, equation (A.9) can be rewritten as

$$P\{AB\} > P\{A\}P\{B\}, \quad (A.10)$$

where $A$ is the event which occurs when the paper $x_1$ illuminated by the light $p_1$ is judged as least dissimilar to the paper $x_2$ illuminated by the light $p_2$; and $B$ is the event which occurs when the paper $x_2$ illuminated by the light $p_2$ is judged as least dissimilar to the paper $x_1$ illuminated by the light $p_1$. As above we have $AB = ((x_1, p_1) \sim (x_2, p_2)) \& ((x_2, p_2) \sim (x_1, p_1)); \bar{A}B = ((x_1, p_1) \sim (x_2, p_2)) \& ((x_2, p_2) \sim (x_1, p_1)); \bar{A} \bar{B} = ((x_1, p_1) \sim (x_2, p_2)) \& ((x_2, p_2) \sim (x_1, p_1)); \bar{A}B = ((x_1, p_1) \sim (x_2, p_2)) \& ((x_2, p_2) \sim (x_1, p_1)).$

Symmetry was tested for each of the 15 illumination conditions in which different lights were involved. (In other words, it was tested only for asymmetric matching.) As in the case of transitivity, the test was performed independently for each observer. The contingency between the events $A$ and $B$ was found to be significant at the level of 0.01 for all the observers and all the illumination conditions apart from one ($p = 0.0745$). More specifically, the null (no symmetry) hypothesis $P\{AB\} = P\{A\}P\{B\}$ was rejected in favour of the alternative (A.10) in all the cases except one.

We conclude that our data testify in favour of both symmetry and transitivity for least dissimilar matching.