Vision Research 49 (2009) 702–707

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Do people match surface reflectance fundamentally differently than they match emitted light?

Jeroen J.M. Granzier, Eli Brenner *, Jeroen B.J. Smeets

Faculty of Human Movement Sciences, VU University, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 20 March 2008
Received in revised form 6 January 2009

Keywords:
Colour vision
Colour matching
Display technology
Methodology

ABSTRACT

We compared matches between colours that were both presented on a computer monitor or both as pieces of paper, with matching the colour of a piece of paper with a colour presented on a computer monitor and vice versa. Performance was specifically poor when setting an image on a computer monitor to match the colour of a piece of paper. This cannot be due to any of the individual judgments because subjects readily selected a matching piece of paper to match another piece of paper and set the image on the monitor to match another image on a monitor. We propose that matching the light reaching the eye and matching surface reflectance are fundamentally different judgments and that subjects can sometimes but not always choose which to match.

1. Introduction

People are very good at determining the contribution of surface reflectance to the light reaching their eyes (Granzier, Brenner, & Smeets, 2009). The lawful relationships underlying this ability obviously do not hold for surfaces that emit light, for which the light reaching the eye can be quite independent of the illumination. Do people consider this? Are there fundamental differences between judgments about surface reflectance and judgments about emitted light? One reason to think that there may be, is that some subjects match patches on a computer screen differently when instructed to make them look identical in hue and saturation than when instructed to make them look as if they were surfaces painted in the same colour (Arend & Reeves, 1986; Cornelissen & Brenner, 1995; see Arend & Spehar, 1993 for similar findings when matching luminance). In those studies, subjects were instructed to treat the patches in different ways. The disadvantage of explicitly instructing subjects to match the colour in a different way is that subjects are encouraged to consider the context in a manner that they probably normally would not, as when asking people to judge the distance to an object in a picture, rather than the distance to the picture itself. Moreover asking subjects to judge the reflectance of simulated surfaces may be particularly confusing because even the stimulus itself is ambiguous (Hurlbert, 1999; Kraft & Brainard, 1999). Several colour vision scientists have therefore designed ingenious systems for presenting surfaces of whatever colour they want within a seemingly natural environment, either by separately illuminating one object (e.g., De Almeida, Fiadeiro, & Nascimento, 2004) or by embedding a monitor screen within a scene in such a manner that it is impossible to see that it is not a real surface (e.g., Hansen, Walter, & Gegenfurtner, 2007; Nascimento, de Almeida, Fiadeiro, & Foster, 2005). Such precautions are taken to ensure that subjects treat the critical surfaces as real surfaces. But do subjects really treat reflected and emitted light differently when such precautions are not taken? Do they switch from evaluating a surface's reflectance to evaluating the composition of the light reflected to the eye when it is apparent that the surface is not real?

We recently found that subjects make better colour matches when real surfaces had to be matched in colour and luminance by selecting the appropriate sample from a colour selector (real coloured papers), than when they had to be matched with a surface on a computer monitor (Granzier, Smeets, & Brenner, 2006). A possible explanation for this is that the simulated surface on the computer monitor was treated fundamentally differently than the real surfaces. However, as mentioned above, the image on the screen is ambiguous when interpreted as a reflecting surface, so the poorer performance may just be a consequence of this ambiguity. If a fundamental distinction is made between light emitted by a monitor and light reflected by a surface, and subjects can judge both independently, subjects should be better at matching the colours of two images on computer screens and at matching the colours of two real surfaces, than at matching a real surface's colour with that of an image on a computer screen. If emitted and reflected lights are treated in the same manner, there need not be a fundamental difference between real and simulated surfaces. However, if this processing is more than just measuring the light reaching the eye, trying to dissociate between reflectance and illumination is likely to introduce additional variability for any match that

* Corresponding author.
E-mail addresses: j_granzier@hotmail.com (J.J.M. Granzier), e.brenner@fbw.vu.nl (E. Brenner).

0042-6989/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2009.01.004

involves images on a computer screen, especially when the surfaces that are to be matched are under different (simulated) illumination. Here we test all four possible combinations of the ways in which reference and matching colours can be presented. In contrast with the above-mentioned studies in which care was taken to prevent subjects from realising that some surfaces emitted light (Hansen et al., 2007; Nascimento et al., 2005) we did our best to ensure that it was always completely clear whether a surface reflected or emitted light. In contrast with our own previous study (Granzier et al., 2006) we used the same illumination for all surfaces. The light reaching the eye from the reference when it was presented on a computer screen was also the same as the light reflected from the reference when it was a piece of paper. This means that matching the light reaching the eye would also match surface reflectance if this were carried out in the same manner for both kinds of reference, so we can expect to find settings within the same region of colour space for all comparisons unless people match reflected and emitted light fundamentally differently.

2. Methods

The experimental room was split into two parts. Further from the subjects was a region in which the 'reference colours' (i.e. the colours that had to be matched) were presented. Nearer to the subjects was a region in which subjects matched the reference colours. The walls of the room were black. The reference colours were either presented as real coloured papers ('reference paper') or as colours on a calibrated computer monitor ('reference monitor'). The reference was either matched by selecting a real surface of an appropriate colour (from a 'colour selector') or else by setting the matching colour on a second calibrated computer monitor ('adjustable computer monitor'). Many colourful common household objects (such as a waste paper basked, a towel and a cup) surrounded the reference monitor. The reference monitor was about 5 m from the subject. When the reference was a piece of paper it was placed about 3.5 m from the subject, between the subject and the monitor, so that the same lamp illuminated the paper and the other objects (see Fig. 1). The reference paper was held in position by a clip of the kind used to hold photographs. The experimenter placed it manually. It was not placed extremely precisely, and the subject's head was not fixed, but subjects were instructed to maintain a head position for which the reference paper more or less occluded the screen of the computer monitor. The small difference in alignment across trials could have some influence on local contrast, perhaps slightly increasing the variability between trials when matching the reference paper. Part of the white borders of the monitor was visible so that the directly surrounding colours were about the same when the reference colour was presented on the monitor as when it was a piece of paper. The dimensions of the reference paper did not correspond precisely with those of the reference colour presented on the computer monitor, and the paper was clearly closer, so that it was always completely clear that it was a real piece of paper. Thus, subjects were always aware of whether the reference colour was being presented as a self-luminous patch (computer monitor) or as a reflecting surface (paper). During presentations in which subjects had to match reference papers the reference monitor was off.

2.1. The reference papers

There were only six reference colours, but subjects were not aware of this. The coloured papers were A4 format (29.6 × 21.1 cm). Under daylight illumination they looked green, pink, purple, light blue, dark blue and white. Under the lamp that we used to illuminate the scene the reference papers reflected light with 1931 CIE_{xyY} coordinates (0.420, 0.486, 5.77 cd/m^2), (0.518, 0.365, 7.25 cd/m^2), (0.452, 0.405, 11.6 cd/m^2), (0.430, 0.416, 11.1 cd/m^2), (0.346, 0.389, 12.1 cd/m^2) and (0.456, 0.417, 14.9 cd/m^2).

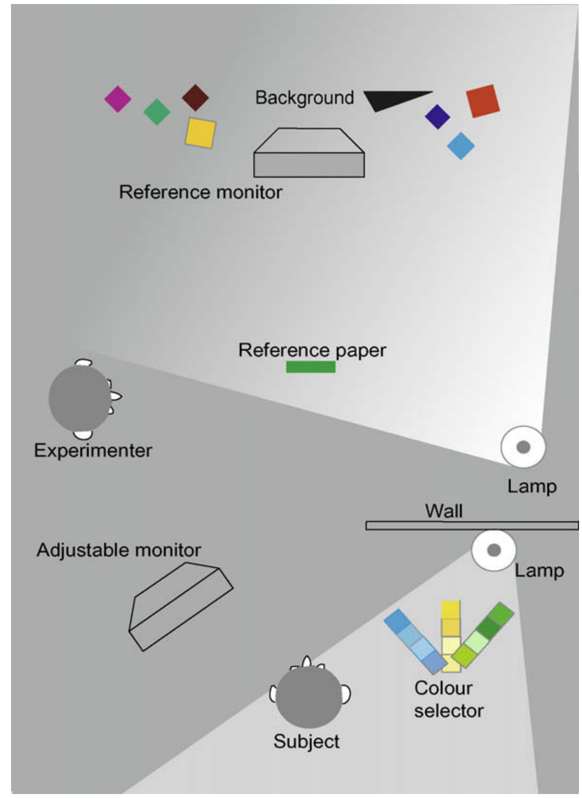

Fig. 1. Schematic overview of the experimental room. Subjects sat 5 m from the reference when it was presented on the computer monitor and 3.5 m from the reference when it was a piece of paper. The reference monitor was embedded in a background of common household objects. A single lamp illuminated the whole scene. Subjects either matched the reference colour with the colour selector or with an adjustable patch on a computer monitor.

2.2. The reference monitor

When the reference was presented on the computer monitor the whole monitor screen was filled with the reference colour. The reference monitor had an effective image size of 32 cm × 23 cm (1280 × 1024 pixels; 85 Hz; 8 bits per gun). The lamp that illuminated the background (and the papers when the reference papers were used) also illuminated the monitor, so we conducted the calibration (using a Minolta CS-100A chroma meter) with this lamp on. This allowed us to select reference colours on the monitor such that the light that reached the subject' eyes was as close as possible to that reflected by the six pieces of paper (see values above; the median error expressed as a distance in CIE_{xy} was 0.01; the median error in luminance was 0.6%). In these cases the light reaching the eyes was a combination of emitted and reflected light. The outer edges of the computer monitor (white plastic) were 4.5 cm wide.

2.3. The colour selector

For colour matching using real papers we used a colour selector (Pantone, New Jersey, 1984). Subjects had to select the sample that best matched the colour and luminance of the reference. Subjects were free to leaf through the pages until they found a suitable sample. Once they had found a good match, subjects read out the number of the matched colour and the experimenter wrote down the number and presented the next reference colour. When matching with the colour selector, the adjustable computer monitor (see below) was off.

2.4. The adjustable computer monitor

For colour matching using emitted light we used a second calibrated computer monitor ($40\ \text{cm} \times 30\ \text{cm}$; 1280×1024 pixels; 90 Hz; 8 bits per gun) that was 1 m from the subject. Subjects had to match the colour and luminance of an adjustable patch to the colour and luminance of the reference. There were only two colours on the screen: the colour set by the subject (within a 5 deg diameter adjustable patch at the centre of the screen) and a uniform background ($10\ \text{cd/m}^2$) with the same CIE_{xy} coordinates (0.47, 0.42) as the light from the "white" background of the colour selector when illuminated by the lamp that was used (see below). The patch's chromaticity was manipulated (within the part of the CIE colour space that we could render on the monitor) by moving the computer mouse. Subjects could manipulate the luminance by pressing the arrow keys on the computer keyboard. They indicated that they were content with the match by pressing the mouse button. The initial hue of the adjustable patch was determined at random for each match from within the range that could be rendered. The luminance of the adjustable patch was $10\ \text{cd/m}^2$ for the first match, but it remained at whatever value the subject set for the next trial.

2.5. Lamps

We used two lamps in our set-up (see Fig. 1). One lamp illuminated the scene (including the reference monitor and reference paper when present). This lamp was always on. Another lamp was only used when subjects matched the reference colour with the colour selector. The two lamps were similar in both intensity and colour. Both lamps had 1931 CIE_{xy} coordinates of (0.47; 0.41) as measured directly with a Minolta CS-100A chroma meter. The lamp illuminating the scene was positioned between the subject and the scene, slightly to the right of the observer (see Fig. 1), so that it would illuminate the front surface of the reference paper as well as the background with the reference monitor. The lamp illuminating the colour selector was positioned in such a way that it only illuminated the colour selector and the black surface on which it rested, but obviously the subjects' hands were visible when manipulating the colour selector.

2.6. Subjects and procedure

One author and three subjects who were naïve as to the purpose of the experiment each took part in the four sessions of the experiment. All had normal colour vision as tested with Ishihara colour plates (Ishihara, 1969). Within each session the subject adapted to the light in the room for 5 min whilst receiving instructions, and then made 30 matches (six reference colours, each presented five times). This was done in a separate session for each of the four comparisons. Within each session, the reference colours were presented in an arbitrary order. Subjects could take as much time to find a suitable match as they liked. The experimenter changed the reference manually after each match had been made. For the reference paper he replaced the paper held by the clip (see above). For the reference monitor he typed a number corresponding to the desired reference colour and the uniform colour on the screen changed accordingly. When the reference was presented on the computer monitor there was obviously no reference paper attached to the clip. The order of conditions in which subjects were tested was counterbalanced (Latin square). The colours presented on the computer monitor were clearly emitted and were not seen as real surfaces, whereas the real papers were clearly identified as being real surfaces. We instructed subjects to match the colour, without any further explanation. By doing so we expected them to automatically judge the reflectance of the papers and the light emitted from the monitor. We examined whether matches are more precise when the two surfaces that are to be matched are both either real surfaces or both emit light (computer monitor). Subjects received no training or feedback.

2.7. Analysis

Since the illumination of the scene was always the same we characterised all surfaces' colours by the 1931 CIE_{xyY} coordinates of the light reaching the eye. We first measured the light that each of the chosen samples of the colour selector reflected when illuminated by the lamp illuminating the colour selector during the experiment. We then determined the mean CIE_{xy} values of each subject's matches for each of the six reference colours in each of the four conditions (reference paper-colour selector, reference monitor-adjustable monitor, reference paper-adjustable monitor and reference monitor-colour selector, where the first term indicates how the reference was presented and the second indicates how the reference was matched). This provided us with 24 mean matches for each condition (four subjects; six reference colours).

Besides simply plotting these matches we also wanted to determine a single value (for each condition) for the variability between subjects. For this we chose the median distance in CIE_{xy} colour space between the four subjects' colour matches for the same reference. These were then averaged across the six reference colours. We also determined an average bias for each subject (for each condition) by averaging the coordinates of the matches for the six reference colours, averaging the coordinates of the reference colours themselves, and determining the distance (in CIE_{xy} colour space) between these two. We report the average amplitude and direction of this bias. Finally, we also determined these measures of variability between subjects and of their average biases after transforming the data into 1976 $\text{CIE}_{u'v'}$ values to examine whether for some reason the results depend critically on the colour space that is used.

3. Results

We examined whether colour matches are worse (larger systematic errors that differ between subjects) when two patches are presented in different ways (matching the colour on a computer monitor to that of a real piece of paper) than when they are both presented in the same way (both as real surfaces or both on a computer monitor). We expected this to be the case if the way we judge colour is fundamentally different for real surfaces (for which reflectance is judged) and for light emitted by a source such as a computer monitor (for which the light itself is judged), because it does not make sense to match emitted light with surface reflectance. In contrast, if colour is always judged in the same way, we expect to find similar matches in all conditions. In the latter case the overall consistency in the matches is expected to depend on how well each individual surface is judged. If images on a screen are particularly difficult to evaluate, then all matches involving such images will be more variable, and matching two images on screens will be the most variable. Since the methods that we use differ in the resolution of the individual matches we concentrate on the variability between subjects and on systematic biases.

On average, when matching with the colour selector, individual subjects chose 2.8 different samples to match the five presentations of each reference (the median distance in CIE_{xy} between any chosen sample and the nearest other chosen sample was 0.006). This indicates that the steps of the colour selector are small enough to obtain a reasonable (average) match, but trying to determine a measure of the variability within the five settings is probably not meaningful (so we do not). Fig. 2 shows the individual average colour matches for each reference in each of the 4 conditions. The crosses show what would be perfect matches of the light reaching the eyes. In general, the average matches are close to the perfect values. However, there are clearly exceptionally large biases in subjects' colour matches for the condition in which reference papers are matched on an adjustable computer monitor. It appears as if subjects came closest to matching the light reaching their eyes in the reference monitor-adjustable monitor matching condition.

To try to quantify these observations we determined the variability between subjects' average matches (for each reference) and the overall systematic bias with respect to matching the light reaching the eyes (for each subject) as explained in Section 2.7. Fig. 3A shows the average variability in colour matches between subjects for each of the four conditions. A repeated measures ANOVA on the variability between subjects (four conditions for each reference colour) revealed a significant difference between the conditions ($F_3 = 22.6$, $p < 0.0001$). Post-hoc tests showed that there was significantly more variability between subjects in the condition in which they matched reference papers with an adjustable computer monitor than in the other three conditions. There were no significant differences between the other three conditions. The values for CIE$_{u'v'}$ (discs) are very similar to those for CIE$_{xy}$ (except for a scaling factor).

Fig. 3B shows the average overall bias in the colour matches for each condition. A repeated measures ANOVA on the magnitude of the bias (four conditions for each subject) revealed a significant difference between the conditions ($F_3 = 9.6$, $p < 0.01$). Post-hoc tests showed that there was a significantly larger bias in the condition in which subjects matched reference papers with an adjustable computer monitor than in the other three conditions. There were no significant differences between the other three conditions. A repeated measures ANOVA on the direction of the bias revealed a significant difference between the conditions ($F_3 = 396, p < 0.0001$). Post-hoc tests showed that the direction of the bias depended on

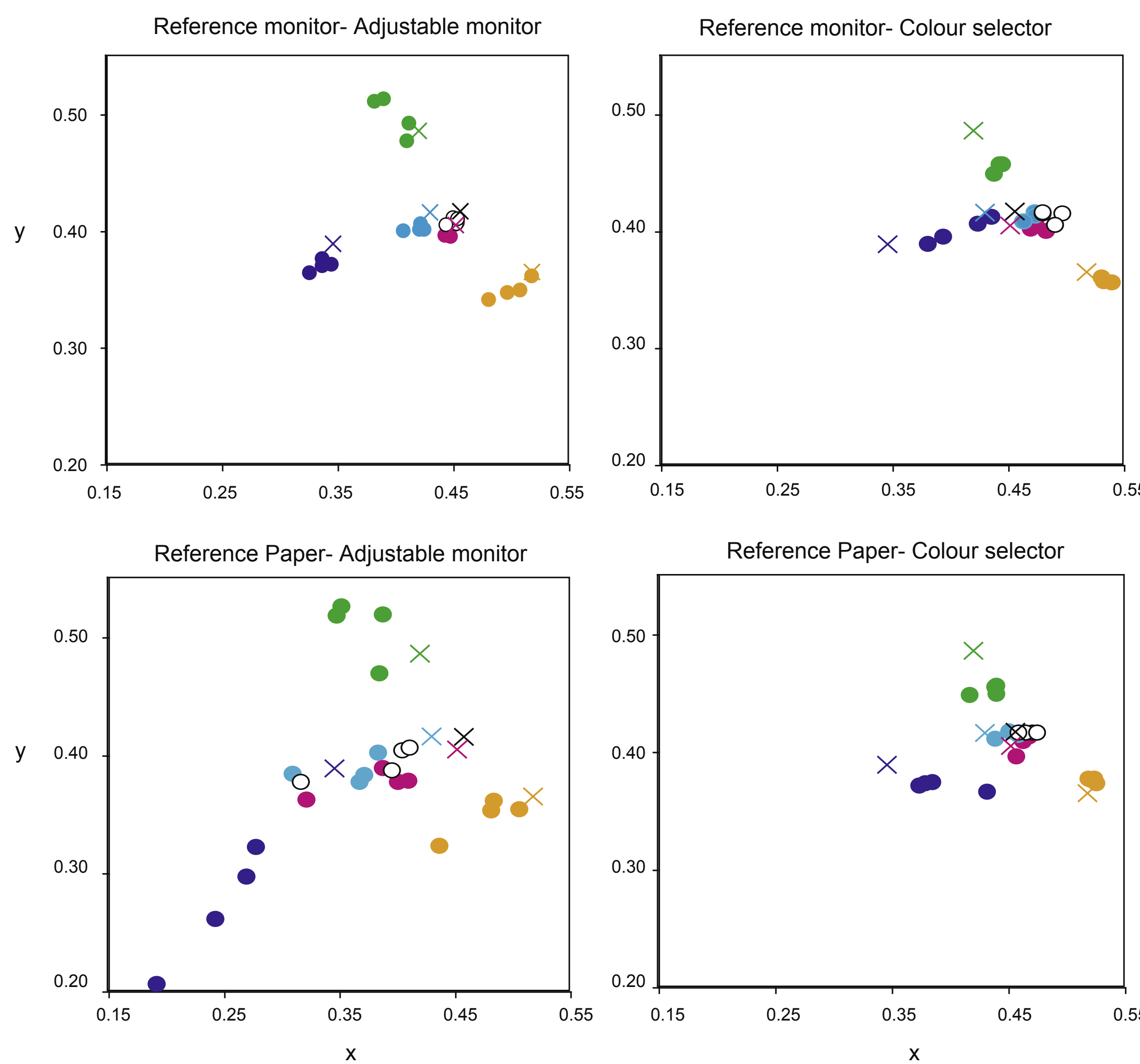

Fig. 2. 1931 CIE coordinates of each of the four subjects' average colour matches (circles) and of the light reaching the subjects' eyes from the reference (crosses) in each of the four conditions (different panels). The different colours of the symbols represent different reference colours. (For interpretation of the references to colour in this figure and the following legend, the reader is referred to the web version of this paper.)

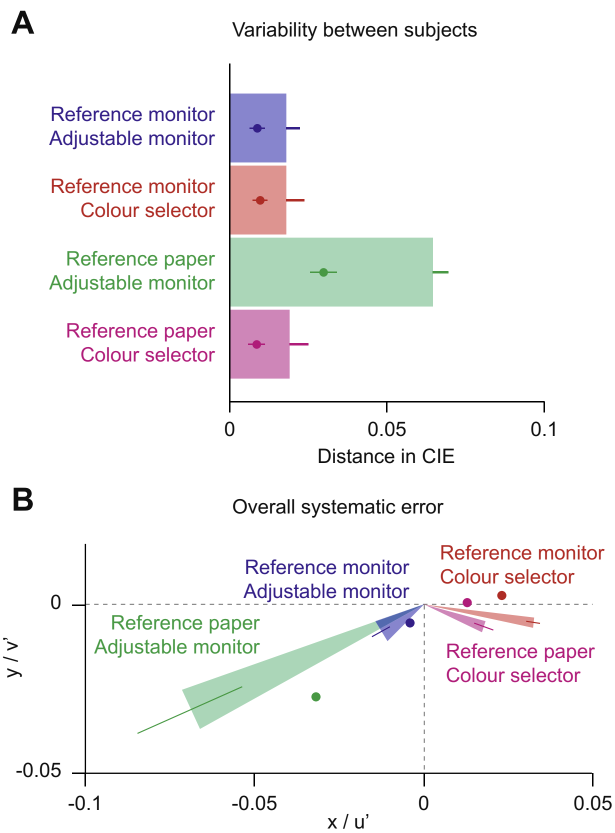

Fig. 3. Two measures of the quality of the match. (A) As a measure of the variability between subjects, we determined the median distance in 1931 CIE_{xy} colour space between the four subjects' (average) colour matches for each reference, and averaged these values across references. (B) As a measure of the overall bias, we determined the amplitude and direction of the mean bias in the (average) matches for the six references, averaged across subjects. The discs indicate the values of the same measures in 1976 $CIE_{u'v'}$. Error bars are standard errors across the six references and four subjects in A and B, respectively. The widths of the 'bars' in B indicate the average ±1 standard error (across subjects) in the direction of the bias.

whether the matching was done with the adjustable computer monitor or with the colour selector. The only comparisons that were not significantly different were the comparison between matching the two kinds of reference with the colour selector and the comparison between matching the two kinds of reference on the adjustable computer monitor. Again the values for $CIE_{u'v'}$ are not fundamentally different.

4. Discussion

Our main finding is that subjects were particularly poor at matching a piece of paper within a scene with an adjustable patch on a computer screen. Since the images reaching the eyes were very similar when the reference surface was a piece of paper and when it was an image on the monitor, a considerable difference in the way they are matched indicates that the visual system treats the two kinds of surfaces differently. Whether the reference was matched on the adjustable monitor or with the colour selector clearly does influence the images that the eyes are exposed to, so it is not too surprising that the direction of the bias was different when subjects matched the reference in a different manner (but not when the reference was presented in a different manner; Fig. 3B). As already mentioned in Section 2, there were some differences between the images on the retina when reference colours were presented on a computer monitor and when they were presented as real papers. The paper was nearer the subject, filled a slightly larger part of the visual field, and was slightly less uniform in luminance because it was often slightly curved. These differences ensured that the subject was always completely aware of whether he or she was dealing with emitted or reflected light. Apparently these differences do not influence the perceived colour very much because similar samples of the colour selector were chosen to match the same light reaching the eye from pieces of paper and from images on a monitor (reference monitor-colour selector and reference paper-colour selector conditions in Figs. 2 and 3). Thus it is surprising that the same two references are not matched in the same way on the adjustable monitor. Why is performance particularly poor when matching a piece of paper with an adjustable patch of light on a monitor? It is poor both in terms of variability between subjects (Fig. 3A) and in terms of the amplitude of the overall systematic error (Fig. 3B). Presumably there is something exceptionally difficult about matching the surfaces in these conditions.

We maintain that the difficulty is related to the way the retinal images are interpreted rather than to the retinal images themselves. If something about matching with the colour selector – such as seeing the hand – made matching with the colour selector better, or something about adjusting the colour on the monitor made matching on the monitor worse, we would expect to see a different pattern in our results. In that case, since the poorest performance is found in the reference paper-adjustable monitor condition, we would expect the best performance when the other two alternatives were used, which is in the reference monitor-colour selector condition. We would expect to find intermediate performance in the other two conditions. This is not what we find. Thus our subjects' performance is inconsistent with any account that assumes that subjects always match colours in the same way. Subjects must have performed the task in a fundamentally different manner in different conditions.

In the introduction we proposed that subjects might judge reflectance for pieces of paper and the light reaching their eyes for screens that emit light. If they had done so we would have found good colour matches when reference colours were presented as real papers and were matched with other real papers, and when reference colours were presented on a screen and were matched by colours on a screen, and poor matches in the other two conditions. The systematic errors were indeed smallest when matching paper with paper or light from a monitor with light from a monitor (Fig. 3B), but performance was exceptionally poor when reference colours presented on paper were matched with colours presented on an adjustable monitor. When matching references with the colour selector subjects did almost as well when the reference was presented on the monitor as when it was presented as a piece of paper. Thus the distinction between judging reflectance and judging the chromaticity of the light reaching the eyes cannot be as straightforward as we proposed.

In our original proposal we assumed that it would be impossible to treat a surface that clearly reflected light as a light source or a surface that clearly emits light as a reflecting surface. Indeed it seems reasonable to assume that the colours of the reference papers and of the papers of the selector can only be considered in terms of reflectance. However the monitor that is embedded within the scene and is illuminated by the lamp that illuminates the scene emits light, but it also reflects light. The fact that subjects readily matched the image on the monitor with the colour selector suggests that they did not always interpret the image on the monitor in terms of the light reaching their eyes. The fact that subjects were not particularly bad at matching colours presented on the reference monitor suggests that they did not try to dissociate reflected from the emitted light, because trying to do so would undoubtedly introduce additional errors. Instead we propose that people can consider reference colours presented on a monitor in terms of either surface reflectance or the light that reaches the eye, but not both simultaneously. When judging surface reflectance the screen is not considered to emit light, so the required distinction between influences of reflection and illumination are the same as one has for the paper reference. When judging emitted light the screen is not considered to reflect light, so subjects attempt to match the light reaching the eye. One may expect the same distinction for the adjustable monitor, but we argue that the variable patch on the adjustable monitor is always considered in terms of the light reaching the eye because the changes in colour prevent one from considering it to be a surface with fixed reflective properties.

If we are correct, subjects match the reference colours presented on the monitor with the colour selector in terms of surface reflectance. They obviously also match real papers with other real papers in terms of surface reflectance. They match the reference colours presented on the monitor with a patch on another monitor in terms of the light reaching the eye. The problem arises when matching a real surface (which is not readily interpreted in terms of the light reaching the eye) with a variable patch on a monitor (which is not readily interpreted in terms of surface reflectance). Thus there is a fundamental difference between judging emitted and reflected light, and subjects are not always free to choose which judgment they make.

Acknowledgments

This work was supported by Grant 051.02.080 of the Cognition Program of the Netherlands Organization for Scientific Research (NWO).

References

Arend, L. E., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A – Optics Image Science, and Vision, 3*, 1743–1751.
Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 1. Illuminance variation. *Perception & Psychophysics, 54*, 446–456.
Cornelissen, F. W., & Brenner, E. (1995). Simultaneous colour constancy revisited: An analysis of viewing strategies. *Vision Research, 35*, 2431–2448.
De Almeida, V. M. N., Fiadeiro, P. T., & Nascimento, S. M. C. (2004). Color constancy by asymmetric color matching with real objects in three-dimensional scenes. *Visual Neuroscience, 21*, 341–345.
Granzier, J. J. M., Smeets, J. B. J., & Brenner, E. (2006). A direct test of the Grey World Hypothesis; a comparison of different matching methods. In *Third European conference on color in graphics, imaging and vision* (pp. 131–136). Leeds, UK.
Granzier, J. J. M., Brenner, E., & Smeets, J. B. J. (2009). Reliable identification by colour under natural conditions. *Journal of Vision, 9*, 1–8, article 39.
Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision, 7*, 1–15.
Hurlbert, A. (1999). Is colour constancy real? *Current Biology, 9*, 558–561.
Ishihara, S. (1969). *Tests for colour blindness: 38 Plate edition*. Tokyo: Kanehara Shuppan Co. Ltd.
Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of colour constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences of the USA, 96*, 307–312.
Nascimento, S. M., de Almeida, V. M., Fiadeiro, P. T., & Foster, D. H. (2005). Effect of scene complexity on colour constancy with real three-dimensional scenes and objects. *Perception, 34*, 947–950.