Perception, 2002, volume 31, pages 151 – 159

DOI:10.1068/p03sp

Scene articulation: dependence of illuminant estimates on number of surfaces

Karina J Linnell
Behavioural Brain Sciences, School of Psychology, University of Birmingham, Birmingham B15 2TT, UK; e-mail: k.j.linnell@bham.ac.uk

David H Foster¶
Computational Neuroscience Group, Department of Optometry and Neuroscience, University of Manchester Institute of Science and Technology, Manchester M60 1QD, UK;
e-mail: d.h.foster@umist.ac.uk

Commissioned for the Special issue on Articulation

Abstract. The ability of observers to detect changes in illuminant over two scenes containing different random samples of reflecting surfaces was determined in an experiment with Mondrian-like patterns containing different numbers of coloured patches. Performance was found to improve as the number of patches increased from 9 to 49. In principle, observers could have used space-average scene colour as the cue ('grey-world' hypothesis) or the colour of the brightest surface in the scene ('bright-is-white' hypothesis), as the two cues generally covary. In a second experiment, observers matched illuminants across different patterns in which the space-average cue and the brightest-patch cue were independently manipulated. The articulation of the patterns was varied: the number of patches increased from 49 (patch width 1 deg visual angle) to over 50 000 (patch width 0.03 deg), while the gamut of colours was held constant. Space-average colour was found to be the dominant cue with all patterns except for those with the largest patches.

1 Introduction

As part of everyday life, we are used to seeing different scenes under different illuminants. Imagine, for example, walking from one room illuminated by daylight to another room illuminated by tungsten light. Provided that each room has enough reflecting surfaces, we should be able to judge whether there has been a change in illuminant without inspecting the source of light directly. How does articulation, here taken as the number of reflecting surfaces in each scene, influence this ability? Clearly, if each scene has only one reflecting surface, it is impossible to disambiguate a difference in illuminant from a difference in the spectral reflectance of the surfaces. But if there are a reasonable number of different surfaces in each scene, then it should be possible to make reliable estimates about the illuminant upon it.

How are such estimates made? One possibility is that the spatial average of all the colours in the scene is used as an estimate of the illuminant colour (the 'grey-world' hypothesis—see eg Buchsbaum 1980; Land 1986; but also Brown 1994). In experiment 1, the ability of observers to detect a change in illuminant over two different random samples of coloured patches forming a Mondrian-like pattern was determined for different numbers of patches and for different changes in illuminants. Performance was found to improve as the number of patches in the patterns increased from 9 to 49. Even so, it is not possible to conclude from such measurements that observers necessarily use space-average scene colour as the cue. Another possibility is that the colour of the brightest patch in the scene is used as the cue, on the grounds that, as the surface apparently reflecting the most light, it is most likely to be white, and therefore most likely to provide an unbiased estimate of illuminant colour (the 'bright-is-white' hypothesis—see eg Land and McCann 1971; McCann et al 1976). The problem is that the colour of the brightest patch in the scene generally covaries with space-average colour.

¶ Author to whom requests for reprints should be addressed.

Accordingly, in experiment 2, the ability of observers to match illuminants across different scenes was measured in Mondrian-like patterns in which the space-average cue and the brightest-patch cue were independently manipulated. The number of patches in each pattern varied from 49 (patch width 1 deg visual angle) to over 50 000 (patch width 0.03 deg). Space-average colour was found to be the dominant cue with all patterns except for those with the largest patches.

2 Experiment 1: Number of colour patches

Observers were presented in temporal sequence with the two Mondrian patterns made up of different, randomly selected surfaces. The first pattern was illuminated by one illuminant and the second pattern by another illuminant, which differed from the first by a variable amount. The illuminants were based on daylights, but manipulated so that their colour varied along and perpendicular to the daylight locus. Observers had to detect the change in illuminant. The patches making up the patterns always had the same size, but their number was allowed to vary.

2.1 *Methods*

2.1.1 *Apparatus.* Stimuli were generated by an RGB colour-graphics system with 8-bit resolution selected on each gun (Ramtek UK Ltd, Hampshire; 4660 series) under the control of a computer (Sun Microsystems Inc, CA, USA; type 3/160) and displayed on a 19-inch RGB monitor (Sony, Japan; Trinitron). Screen resolution was 1280×1024 pixels. The system was calibrated before the experiments, and at intervals during them, with the aid of a photometer and telespectroradiometer whose calibrations could be traced to the National Physical Laboratory.

2.1.2 *Stimuli.* Stimuli were computer simulations of illuminated, square, Mondrian-like patterns presented in a black field. The individual colour patches physically comprising the Mondrian patterns were square and subtended on each side 0.86 deg visual angle. There were 9, 25, or 49 patches in each pattern, which therefore subtended on each side 2.6, 4.3, or 6.0 deg, respectively. Surfaces were drawn randomly from the 1976 *Munsell Book of Color* (Munsell Color Corporation 1976), and their spectral reflectances were taken from a basis-function decomposition (Parkkinen et al 1989), which was used for computational speed and has no theoretical implications here.

The illuminants were all formed from combinations of the basis functions derived from daylights (Judd et al 1964). According to the choice of coefficients of these basis functions, illuminants could be made to shift in colour in any direction in colour space. In terms of the CIE 1976 (u', v') chromaticity diagram, illuminants were shifted from a whitish origin, at $u' = 0.20$ on the daylight locus, to a range of points radiating out from this origin along four different colour directions: 'blue', 'orange', 'green', and 'magenta'. More precisely, blue and orange shifts were from the whitish origin at $u' = 0.20$ to various points along the tangent to the daylight locus at $u' = 0.20$, ranging to $u' = 0.18$ and to $u' = 0.22$, respectively; green and magenta shifts were from the origin to equivalently spaced points along the normal to the daylight locus at $u' = 0.20$. These changes in illuminant colour are modest: thus, u' values of 0.20 and 0.22 on the daylight locus correspond to correlated colour temperatures of about 6100 K and 4200 K, respectively.

2.1.3 *Procedure.* On each trial, one Mondrian pattern was presented for 1 s and it was immediately replaced by a second Mondrian pattern, which was also presented for 1 s. Observers were asked to decide which of two events accounted for the change in the two patterns: either a change in the random sample of reflecting surfaces and their geometry, or a change in the random sample of reflecting surfaces and their geometry in conjunction with a change in the illuminant. The two events were equally likely.

When a change in illuminant occurred, its magnitude took one of five different values represented by five different Euclidean distances in (u', v') space: 0.0066, 0.0131, 0.0197, 0.0263, and 0.0438, in any one of the four different colour directions. An experimental session consisted of eight blocks of 80 trials. Within sessions, the number of coloured patches from which each pattern was built remained the same, but across sessions it varied. In all, observers completed three sessions with each of the numbers of patches.

2.1.4 *Observers.* There were two observers, GP and TH. Each had normal colour vision as assessed with the Farnsworth – Munsell 100-Hue test and normal Snellen acuity. Each was unaware of the purpose of the experiment.

2.2 *Results*

Detection performance was quantified in terms of the discrimination index d' from signal-detection theory (Green and Swets 1966). Figure 1 shows, for each of the two observers, values of d' plotted against number of patches for each of the four colour directions of illuminant shift. The different symbols, joined by straight lines, correspond to different sizes of illuminant shift. Performance that was equal to chance or limited by the maximum values of d' computationally possible is indicated by the horizontal grey lines.[(1)]

2.3 *Comment*

Observers' ability to detect a change in illuminant over two different scenes improved as the number of patches in the scenes increased. Within the dynamic range available, this rule held whether the change in illuminant was small or large, but levels of performance were slightly better for displacements in the direction perpendicular to the daylight locus (green and magenta) than along the direction parallel to the daylight locus (blue and orange).

To determine whether the data of figure 1 are consistent with the assumption of illuminant estimates based on space averaging, a calculation was made of the performance of an ideal observer who could perfectly extract space-average colour. This ideal observer was presented with exactly the same stimuli as one of the actual observers (TH). The first step to the calculation was to derive the space-average colour of each Mondrian pattern. If the Euclidian distance between the space-average colours of any two sequentially presented Mondrian patterns was greater than some criterion value, the ideal observer responded that an illuminant change had occurred, and if it was less than this criterion value the ideal observer responded that no illuminant change had occurred. Criterion values were chosen so that the false-alarm rates of the ideal observer were the same as those of the actual observer. The d' values calculated from these hypothetical responses for the three sizes of Mondrian patterns and five sizes of illuminant shifts were broadly similar to the d' values that observer TH actually obtained.

Even so, this correspondence does not establish that space-average colour is the cue. As indicated in section 1, the colour of the brightest patch in the pattern generally covaries with the space-average colour. In analyses of experiment 1, it was found

[(1)] If HR is the hit rate, FAR the false-alarm rate, and z the inverse of the cumulative unit normal distribution, then $d' = z(\mathrm{HR}) - z(\mathrm{FAR})$. With a finite number of trials, it is possible for either HR or FAR to be 0 or 1, with the result that d' is undefined. One way to deal with this problem is to replace the scores of 0 and 1 by $1/(2n)$ and $1 - 1/(2n)$, respectively, where n is the number of trials involved (compare eg Cox 1970). This correction sets upper and lower limits on the empirical values of d'.

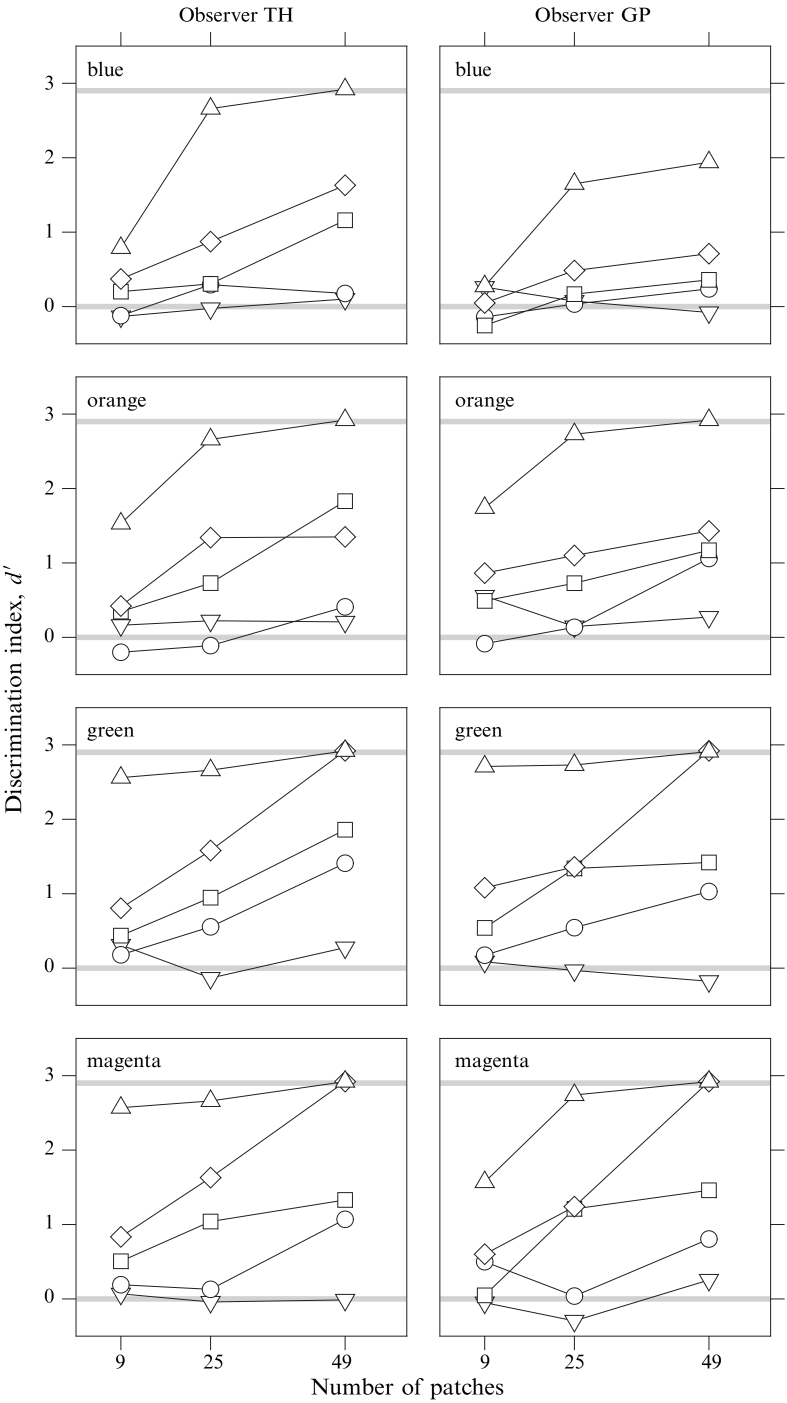

Figure 1. Detectability of a difference in illuminants across two, different, sequentially presented Mondrian patterns. Discrimination index d' from signal-detection theory (Green and Swets 1966) is plotted against the number of patches in each of the two patterns for each of four colour directions of illuminant shift (indicated in each panel). The different symbols, joined by straight lines, correspond to different sizes of illuminant shift [Euclidean distances in (u', v') space: inverted-triangle 0.0066, circle 0.0131, square 0.0197, diamond 0.0263, and triangle 0.0438]. Performance equal to chance or limited by the maximum values of d' computationally possible is indicated by the horizontal grey lines. Data for two observers.

that the two observers' hit and false-alarm rates were predicted not only by shifts in space-average colour but also by shifts in the colours of the patches with the highest luminances. To determine whether illuminant information can actually be derived from estimates of space-average colour alone, it is necessary to manipulate this cue independently of the cue provided by the brightest patch. This was done in experiment 2.

3 Experiment 2: Space-average versus brightest-patch cue

The cue provided by space-average colour was independently biased by building 'dual-cue' Mondrian patterns from a fixed sample of 49 Munsell surfaces which, on average, reflected more light in one (orange) region of the visible spectrum than in any other region. One of the 49 surfaces was, however, a whitish surface that reflected light at all wavelengths approximately equally, and had the highest luminance under all the illuminants used to illuminate the dual-cue pattern. Observers' judgments about illuminant colour would therefore depend critically on whether they used space-average colour or the colour of the brightest patch. A second 'match' Mondrian pattern was built from a fixed unbiased sample of 49 Munsell surfaces, for which both space-average colour and the colour of the highest-luminance patch were good predictors of illuminant colour. The observers' task was to adjust the colour of the illuminant on this match Mondrian pattern until the illuminant was judged the same as that illuminating the dual-cue Mondrian pattern. The two patterns were presented side-by-side and haploscopically, so that they could be compared simultaneously and with minimum interaction.

Although the dimensions of the individual patches were the same in the match and dual-cue patterns, they varied across experimental conditions. To keep pattern size constant, the number of individual patches generated from each of the 49 Munsell spectral reflectances in each pattern was increased as patch size decreased, from 49 with 1.0-deg (32-pixel) patches to over 50 000 with 0.03-deg (1-pixel) patches. With the last, it was impossible to identify individual colours, and observers were expected to use the space-average cue. But, as patch size increased, the colour of the highest-luminance patch would have been more easily identified, and more likely to provide a competing cue. Because the 49 Munsell spectral reflectances of the dual-cue Mondrian pattern and the 49 of the match pattern were each fixed, these variations in articulation were independent of changes in colour gamut.

3.1 *Methods*

3.1.1 *Apparatus.* The apparatus was the same as in experiment 1.

3.1.2 *Stimuli.* As in experiment 1, stimuli were computer simulations of Mondrian patterns of illuminated Munsell surfaces. The simulations relied on the same basis-function descriptions of spectral reflectances and illuminants as in experiment 1. Two square Mondrian patterns, each of side 7.0 deg were presented side-by-side on a black background with their inside vertical edges separated by 3.0 deg. The match Mondrian pattern, presented on the left side of the display, and the dual-cue Mondrian pattern, presented on the right side, had equal numbers of patches. Across different conditions of the experiment, patch size varied from side 1 pixel (0.03 deg), through 2, 4, 8, and 16 pixels, to side 32 pixels (1.0 deg), but patch colour was always equally likely to be the colour of one of the 49 Munsell surfaces associated with each pattern. In Mondrian patterns with patches of side 32 pixels, each of the 49 Munsell surfaces associated with the pattern occurred only once, whereas in Mondrian patterns with side 1-pixel patches, each of the 49 surfaces occurred 1024 times.

The 49 Munsell surfaces associated with the match pattern reflected on average the same amount of light in all regions of the visible spectrum, whereas those associated with the dual-cue pattern reflected on average more light in the orange region of the visible spectrum. One of the 49 surfaces associated with both match and dual-cue patterns, however, reflected light at all wavelengths approximately equally, and was the brightest surface under all the coloured illuminants on the dual-cue pattern.

Across trials, the dual-cue Mondrian pattern was presented under eight different illuminants: two blue, two orange, two green, and two magenta. Like the illuminants in experiment 1, they all fell along either the tangent or the normal to the daylight locus at a whitish origin $u' = 0.20$ in (u', v') space. The two blue illuminants fell on the tangent to the daylight locus and had u' values of 0.18 (more saturated) and 0.19 (less saturated). The two orange illuminants also fell on the tangent to the daylight locus but had u' values of 0.22 (more saturated) and 0.21 (less saturated). The two green and the two magenta illuminants fell at equivalently spaced points on the normal to the daylight locus, each side of the point $u' = 0.20$. The (u', v') coordinates of all of these illuminants, except the more saturated orange one, are plotted as the colours of the highest-luminance patch in the dual-cue pattern in each of the graphs in figure 2 (triangles).

3.1.3 *Procedure.* In each trial, the match and dual-cue Mondrian patterns were presented continuously. Observers were asked to adjust the colour of the illuminant on the match Mondrian pattern (the illuminant being set initially to be whitish) until it was judged to be the same as the illuminant on the dual-cue Mondrian pattern. While making these adjustments, observers were encouraged to switch attention from one pattern to the other. Within each experimental session, patch-size (and therefore patch number) was blocked and randomised, and within each patch-size block observers made eight matches, one for each of the eight illuminants on the dual-cue patterns. There were six experimental sessions in all, so that observers repeated matches for each combination of illuminant and patch size six times. Different random arrangements of patches were used in trials with different patch sizes and in different sessions.

3.1.4 *Observers.* There were two observers, GP and ME (one of whom participated in experiment 1). Each had normal colour vision as assessed with the Farnsworth – Munsell 100-Hue test and normal Snellen acuity. Each was unaware of the purpose of the experiment.

3.2 *Results*

Because the match pattern was, on average, neutral, its space-average colour after the matching could be used to represent observers' estimates of the illuminant on the dual-cue pattern. Figure 2 shows average illuminant estimates (filled circles) by each of the two observers for each of seven of the eight illuminants on the dual-cue patterns with three of the six patch sizes.[2] For comparison, straight lines connect the two theoretical estimates: squares for illuminants estimated by the space-average colour of the dual-cue pattern and triangles for illuminants estimated by the colour of the highest-luminance patch in the dual-cue pattern.

[2] Data are not presented for the more saturated of the two orange illuminants on the dual-cue patterns because, when the illuminant on the match patterns was judged the same colour to observers, it was so saturated that the colour of the reflected light from some surfaces could not be reproduced by the monitor.

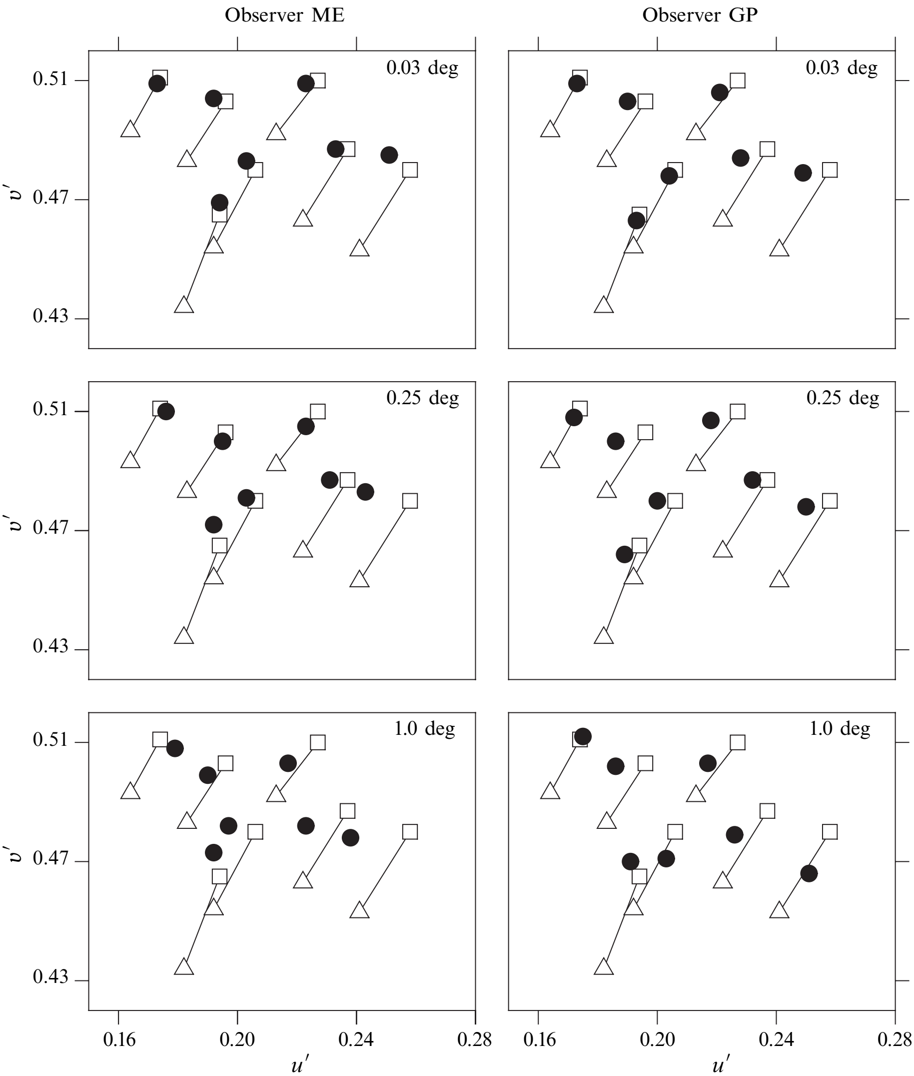

Figure 2. Illuminant estimates for three different patch sizes. The average match (filled circles) made by each observer is plotted in the CIE 1976 (u', v') chromaticity diagram for each of seven of eight illuminants on dual-cue Mondrian patterns and three of the six patch sizes used (indicated in each panel). For comparison, straight lines connect the two theoretical estimates: squares for illuminants estimated by the space-average colour of the dual-cue pattern and triangles for illuminants estimated by the colour of the brightest patch in the dual-cue pattern. Data for two observers.

3.3 *Comment*

Observers set the colour of the illuminant on the match pattern so that its space-average colour was close to the space-average colour of the dual-cue pattern for almost all sizes (and numbers) of patches making up the patterns, from side 0.03 deg to side 0.5 deg. Only for the largest patches (1-deg side) were matches shifted towards those estimated by the colour of the highest-luminance patch. Even then, the shift was only partial, and observers still seemed to use space-average colour, albeit less strongly than when the patches were smaller. Not only does space-average information appear to be strongly preferred by observers, it also appears to be extracted remarkably reliably: performance altered little as patch size increased from side 0.03 deg to side 0.25 deg.

4 Summary and conclusions

Modest changes in illuminant on different scenes could be reliably detected when there were between 9 and 25 patches in the stimulus patterns. As experiment 1 showed, within the dynamic range set by the size of the illuminant changes, observers' performance improved as the number of patches increased. This performance was similar to that shown by an ideal observer using space-average scene colour as the cue to illuminant colour. Yet, both in these and in naturally occurring scenes, space-average colour and the colour of the brightest region covary, and, in principle, observers could have used either cue. In experiment 2, these two cues were pitted against each other. The number of patches in the patterns was varied while the gamut of colours remained constant. Space-average colour remained the dominant cue with patterns in which the constituent patches ranged in size from 0.03 deg to 0.5 deg across, and was still a strong cue with patches 1.0 deg across.

Beck (1959, 1961) addressed this question of how the illumination of a surface is judged, but using grey-level materials, and, in one experiment, wood surfaces. He concluded that for patterns in which there are no clearly identifiable areas of high brightness, judgments about illuminant are strongly influenced by space-average luminance, whereas for patterns in which there are clearly discriminable patches of high brightness, judgments are strongly influenced by the luminance of those patches (Beck 1961). Although no strong effect was found here in the chromatic domain, it seems likely that if patch size had been increased beyond 1 deg visual angle, and the number of coloured surfaces decreased, the cue offered by the colour of the brightest patch could have become more influential. In a rather different task, the appearance of 'colour tautomi' arrays of just five colour patches has been shown to depend strongly on the presence of a white in the array (McCann 1992). It is also possible that, if the denser and more richly sampled Mondrian patterns used here had contained shape-from-shading information and the whitish patches were interpreted as highlights, observers might have attached greater significance to these local cues, and correspondingly less to space-average ones (for tests of these cues, see Hurlbert et al 1990; Yang and Maloney 2000). Nevertheless, for the flat stimuli used here, space-average colour seems to be the decisive cue for judging the illuminant. It remains for future work to distinguish between a true space average and related spatial measures such as the median or logarithmic average.

Acknowledgements. This work was supported by the Wellcome Trust and the Biotechnology and Biological Sciences Research Council. We thank K Amano, J J McCann, S M C Nascimento, and S Westland for helpful discussion, and J P S Parkkinen, J Hallikainen, and T Jaaskelainen for providing their data for the spectral reflectances of the Munsell surfaces. Some of the material described here was reported in Linnell and Foster (1997).

References

Brown R O, 1994 "The world is not grey" *Investigative Ophthalmology & Visual Science* **35** 2165

Buchsbaum G, 1980 "A spatial processor model for object colour perception" *Journal of the Franklin Institute* **310** 1–26

Cox D R, 1970 *The Analysis of Binary Data* (London: Methuen)

Green D M, Swets J A, 1966 *Signal Detection Theory and Psychophysics* (New York: John Wiley)

Hurlbert A C, Lee H-C, Bülthoff H H, 1990 "Specularities as cues to illuminant color" *Perception* **19** 333

Judd D B, MacAdam D L, Wyszecki G, 1964 "Spectral distribution of typical daylight as a function of correlated color temperature" *Journal of the Optical Society of America* **54** 1031–1040

Land E H, 1986 "Recent advances in retinex theory" *Vision Research* **26** 7–21

Land E H, McCann J J, 1971 "Lightness and retinex theory" *Journal of the Optical Society of America* **61** 1–11

Linnell K J, Foster D H, 1997 "Space-average scene colour used to extract illuminant information", in *John Dalton's Colour Vision Legacy* Eds C Dickinson, I Murray, D Carden (London: Taylor and Francis) pp 501 – 509

McCann J J, 1992 "Rules for colour constancy" *Ophthalmic and Physiological Optics* **12** 175 – 177

McCann J J, McKee S P, Taylor T H, 1976 "Quantitative studies in Retinex theory. A comparison between theoretical predictions and observer responses to the 'Color Mondrian' experiments" *Vision Research* **16** 445 – 458

Munsell Color Corporation, 1976 *Munsell Book of Color — Matte Finish Collection* (Baltimore, MD: Munsell Color Corporation)

Parkkinen J P S, Hallikainen J, Jaaskelainen T, 1989 "Characteristic spectra of Munsell colors" *Journal of the Optical Society of America A* **6** 318 – 322

Yang J N, Maloney L T, 2000 "Surface color and specularity: Testing the D'Zmura – Lennie – Lee Model" *Investigative Ophthalmology & Visual Science* **41** 1255 – B630