Author Manuscript

Published in final edited form in *Vision Research*:
Craven, B. J. and Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research*, **32**, 1359-1366; doi:10.1016/0042-6989(92)90228-B

An Operational Approach to Colour Constancy[1]

B. J. Craven[2] and David H. Foster[3]

Department of Communication and Neuroscience, Keele University, Staffordshire ST5 5BG, England

Abstract

Colour constancy is traditionally defined as the invariance of perceived surface colours under changes in the spectral composition of the illuminant. Existing quantitative studies show that, by this definition, human subjects show poor colour constancy. A different and complementary aspect of colour constancy is considered which is concerned with the ability of a subject to attribute correctly changes in the colour appearance of a scene either to changes in reflecting properties of the surfaces that make up the scene, or to changes in the spectral composition of the illuminant. Data are presented showing that, if the changes in the appearance of a scene were sufficiently great, subjects were capable of making the required discriminations highly reliably, and without scrutiny.

Keywords

Colour vision; colour constancy; Munsell papers; daylight; colour discrimination; perceptual invariants; surface spectral reflectance

INTRODUCTION

The term *colour constancy* is generally taken to refer to the invariant appearance of surface colours under changes in the spectral composition of the illuminant. Hering (1920) reported observations of colour constancy, and the phenomenon was considered by Helmholtz (1911), but the most well-known demonstrations of the phenomenon are the colour Mondrian experiments of Land (1977). In these demonstrations, Land claimed to show that two surfaces that reflected lights of different chromaticity could appear identical in colour if the spectral reflectance functions of the surfaces were the same. He concluded that the perceived colour of a surface corresponds to its spectral reflectance and not necessarily to the composition of the reflected light, and that this compensation for the illuminant is achieved at a low level in the visual system.

[1] Portions of this work were reported at the 13th European Conference on Visual Perception, Paris 1990 [Craven B. J., Foster, D. H. & Sale, E. R. H. (1990) *Perception*, 19, 333]

[2] Present address: Department of Product Design Engineering, Glasgow School of Art, 167 Renfrew Street, Glasgow G3 6RQ, UK; email: ben@bencraven.org.uk

[3] Present address: School of Electrical and Electronic Engineering, University of Manchester, Sackville Street, Manchester M13 9PL, UK; email: d.h.foster@manchester.ac.uk

Quantitative techniques have been developed to allow estimates to be made of the extent or completeness of colour constancy. McCann, McKee and Taylor (1976) demonstrated that subjects show a form of colour constancy with Mondrian displays, their results being expressed in terms of the integrated reflectances of matching Munsell chips. Subsequently, the use of colour CRT displays has allowed more easily interpreted measurements of colour constancy to be made. These measurements have typically required the subject to make matches between (simulated) surfaces illuminated by different (simulated) illuminants. In one study (Arend & Reeves, 1986; see also Reeves, Arend & Schirillo, 1989) a task was used in which subjects were shown simulations of two identical Mondrians of Munsell papers under different illuminants. Subjects were required to adjust the colour of a specified patch in one array until it appeared to match the corresponding patch in the other array. The settings of a perfectly colour-constant subject and of a completely non-colour-constant subject in this experiment plot at predictable locations in the CIE chromaticity diagram, and thus it is possible to observe how colour constant a subject is by the positions of his or her settings relative to these extremes. Two of the three subjects achieved approximate colour constancy when they were explicitly instructed to match the colours of the papers rather than the reflected lights; these two subjects, however, had knowledge of the colours that the papers should have taken on under the different illuminants; the third, naive subject showed weak colour constancy. All three subjects showed very weak colour constancy if they were instructed to match the colours of lights rather than papers. Similar results have been reported by other investigators using similar techniques (Tiplitz Blackwell & Buchsbaum, 1988; Valberg & Lange-Malecki, 1990). These results are evidence against the notion of colour constancy as an automatic, low-level, and complete compensation for changes in the composition of an illuminant.

There is a need to reconcile the apparently convincing everyday experience of colour constancy with the rather approximate colour constancy reported in the preceding experiments. Arend and Reeves (1986) briefly considered that colour constancy might involve a high-level awareness of a change in illuminant which enables the subject to disregard changes in apparent surface colour. This paper makes this view of colour constancy more general, and introduces experimental techniques by which it might be investigated.

Consider the following simple experiment: if one switches on an incandescent lamp in a room already illuminated by daylight, the colours of objects in that room appear to change and to stay changed (provided, of course, that the daylight is not sufficiently intense to render changes due to the lamp sub-threshold). This observation appears to conflict with the traditional definition of colour constancy, which implies that the apparent colours of objects should remain constant. Despite the change in appearance we do not generally infer from our observation that the physical properties of the surfaces in the room have changed. Instead, we attribute the changes we see to a change in the colour of the incident light.

There are analogies in other visual tasks. Some spatial transformations of the image of an object (e.g. rotations, dilatations) can be accounted for in terms of object movement alone, whereas others necessarily imply deformation of the object. The transformations compatible with a rigid object can be just as perceptible as those requiring deformation of the object. In a similar way, a scene can undergo transformations in some colour space. Some of these transformations can be accounted for in terms of a simple change in illuminant, others necessarily imply a change in the reflecting properties of the scene. The transformations compatible with unchanged reflectances could be just as perceptible as those caused by changing reflectances. The important practical point for a visual system is that these different kinds of transformations should be distinguishable, in the way that they usually are for spatial transformations of an object.

With these considerations in mind we might consider an alternative and complementary property of colour constancy, as follows:

> The ability of a subject to correctly attribute changes in the colour appearance of a scene either to changes in the spectral composition of the illuminant or to changes in the reflecting properties of that scene.

This aspect of colour constancy is not concerned with the nature or extent of any changes in colour appearance, but simply with the subject's interpretation of them. For a certain area of a scene to be identified by a subject as physically unchanged under a change in illuminant, it is not necessary that the area generate some invariant local percept; all that is needed is that it is perceived to stand in relation to other areas in the scene in the same way. An extreme version of this view, not that advocated here, is that one could identify a given surface as being unchanged under a change in illuminant without being able to identify the colour of the surface.

Theoretical considerations lead to the conclusion that, in general, successful colour constancy depends on constraints on the spectral composition of the possible illuminants and on surface spectral reflectance functions (e.g. Maloney, 1986). Clearly, if reflectances and illuminances could be chosen arbitrarily at each wavelength, colour constancy would be impossible. We make no assertions here about what these constraints might be, other than that the constraints obeyed by real surfaces and natural illuminants are likely to be those to which the visual system is accustomed. It is therefore important to use realistic illuminants and reflectances in experiments on human colour constancy.

EXPERIMENTS

Experiment l determined whether subjects were capable of correctly attributing changes in the colours of a scene to changes in the illuminant or to changes in the materials of which the scene was made.[4] Subjects were presented with two successive Mondrian patterns and were required to decide whether they could be related by a change in illuminant or whether a change in material had occurred. As explained later, the two kinds of changes were equivalent in magnitude. To provide a general reference measure of sensitivity to the stimulus changes involved, Experiment 2 made measurements of subjects' performances in detecting changes in a scene, without requiring classification of the changes. The two experiments, the one requiring discrimination, the other detection, used the same stimuli and differed only in details of procedure. Experiment 3 provided a control on Experiment 1, testing for a possible contribution to subjects' performances of a global, illuminant-change cue based on shifts in average Mondrian colour. In the Appendix an analysis is presented which models the decision processes involved in Experiments 1 and 2, allowing the relationship between the results of those experiments to be more clearly appreciated.

Subjects

There were three subjects, all of whom had normal or corrected-to-normal visual acuity and who had been assessed as colour normal on the Farnsworth-Munsell 100-Hue test. Subjects BJC and DHF were aware of the nature and purpose of the experiment, and subject PRG was not.

Apparatus

Stimuli were generated by an RGB colour graphics system with 8-bit resolution on each gun (Ramtek U.K. Ltd, 4660 series) under the control of a computer (Sun Microsystems, Inc. U.S.A., type 31160). Calculation of the chromaticities and luminances required for each stimulus was performed by the computer in the inter-trial intervals. Once loaded into the colour graphics system, the program of instructions required to display the stimulus sequence was executed without further intervention by the host computer; thus the timing uncertainties inherent in a multi-tasking system were avoided. The stimuli were displayed on a 19" high-resolution RGB monitor (Manitron Displays Ltd, U.K., type CLR2005). Screen resolution was 1280 ×1024 pixels. The system was calibrated before the experiments, and at intervals during them. The luminance output of each phosphor was measured with a photometer (LMT GmbH, Berlin, type L1003) as a function of the input value; the phosphor chromaticity coordinates were measured with a telespectroradiometer (Bentham Instruments Ltd, U.K., 200 series). The calibrations of the photometer and telespectroradiometer were in turn verified against incandescent lamps and narrow-band discharge sources respectively, the calibrations of which were traceable to the U.K. National Physical Laboratory.

Stimuli

Displays simulated the illumination of a Mondrian display of Munsell papers illuminated by different phases of daylight. (Any subsequent references to Mondrians, Munsell papers or daylights should be taken to refer

[4] In practice, of course, changes in the colour of materials would often be associated with other cues, including, for example, changes in texture and specularity. Only colour cues were allowed here.

to the corresponding simulations.) Data for the spectral reflectance functions of chips in the *Munsell book of color—matte finish collection* (Munsell Color, Baltimore, Md, 1976) were taken from the principal components analysis by Parkkinen, Hallikainen and Jaaskelainen (1989; personal communication). Using their eight published eigenvectors, Parkkinen *et al.* state that the spectral reflectance functions can be reconstructed such that the average error in CIE *x*- or *y*-coordinates is about 0.001. Maloney (1986) performed a similar analysis on a more restricted set of Munsell colour samples, and also on data for the spectral reflectance functions of natural surfaces collected by Krinov (1947). Maloney reported that as few as six characteristic vectors provided essentially perfect fits to the surface spectral reflectance functions in both data sets. Further, the basis elements of the Munsell set provided excellent fits to the Krinov data. It may be concluded that the eight eigenvectors that were used in the current study to represent spectral reflectances were adequate, and that the Munsell samples provide a realistic substitute for the spectral reflectance functions that might be encountered in the real world. Data for the spectral energy distributions of daylight were taken from the principal components analysis by Judd, MacAdam and Wyszecki (1964). Natural daylights fall about a line on the CIE chromaticity diagram, slightly to the green side of the Planckian locus, and thus can be labelled (but not specified) by a single coordinate. Henceforth daylights are labelled by their 1931 CIE *x*-coordinate. The data of Judd *et al.* (1964) allowed the generation of daylights falling on this locus anywhere between $x = 0.25$ and 0.37 (correlated colour temperatures from about 30,000 to 4400 K). The phosphors on the colour monitor were such that 1149 of the available 1257 Munsell papers could be displayed, each illuminated by any daylight. The 108 non-displayable colours were generally those of highest saturation, particularly in the blue-green region. The luminance of the stimulus displays was on average 4 cd m^{-2}, but individual patches of the displays varied widely about this value. If the stimuli had been composed of real Munsell papers, the equivalent illuminance would have been approximately 30 lx.

The stimulus displays were square with side 6° at the viewing distance of 135 cm, and consisted of an array of 49 rectangular patches. The centres of the patches were arranged on a square grid. Their aspect ratio and size were chosen randomly, within the constraint that no patch should be smaller than 0.86° square, or larger than 1.29° square. The ragged edges thus produced were removed by displaying only those parts of the pattern that fell within the 6° square. The patches of which the pattern was composed were selected at random from the set of 1149 displayable Munsell papers. Each subject performed about 450 trials, in the course of which about 22,000 Munsell papers were selected, with replacement, from the set of 1149 displayable papers.

The two patterns which comprised each stimulus were presented sequentially. This procedure prevented subjects from performing the task by alternating inspection of the two displays, and also eliminated the need for the subject to assume two different illuminants present in the same scene at the same time. To avoid introducing a memory component into the task, the temporal interval between the patterns was zero. In natural viewing conditions, rapid changes in illuminant can occur, for example, when the sun passes behind a cloud or when an incandescent lamp is switched on in a previously daylit room.

The stimulus on each trial lasted for 2 sec. For the first half of this period, a Mondrian was displayed, illuminated by daylight with CIE *x*-coordinate 0.31. It was then replaced for the second half of the exposure (without a dark interval) by a second Mondrian which was derived from the first in one of two ways: (1) the illuminant for all the patches in the pattern was changed by a shift along the daylight locus in either the positive or the negative *x*-direction; or (2) for half of the patches in the pattern, selected at random, the illuminant was shifted along the daylight locus in the positive direction, and, for the remaining half of the patches, the illuminant was given an equal shift in the negative direction. The changes in the Mondrians in both of these conditions are inherently ambiguous: changes in surface spectral reflectance, illuminant composition, or both, could have occurred. The change in the stimulus that occurs with manipulation (1) has the parsimonious interpretation that it is due to a spatially uniform shift in illuminant, and we refer to this operation as a *change of illuminant*. The change that occurs with manipulation (2) could not be obtained by a spatially uniform (or spatially slowly varying) change in illuminant, without a simultaneous change in the spectral reflectances of the patches making up the display. Because non-uniform illuminant shifts of this kind are uncommon, the most parsimonious interpretation of manipulation (2) is of a *change of material*.

The cue to the subject on any trial was determined by the shift in the CIE *x*-coordinate of the illuminant, whether the shift was applied with uniform sign (illuminant-change trials) or with random sign (material-change trials). The magnitude of the shift is denoted by Δx. On each trial, the differences between the "initial" and "final" patterns, considered on a patch-by-patch basis, were thus of the same magnitude whether it was an illuminant-change trial or a material-change trial. Therefore, the subject would not have been able to make the discrimination correctly on the basis of observation of a single patch; to perform the task information about colour changes in multiple patches of the display had to be combined. It should be noted that there is no simple and precise relationship between the initial chromaticity of a given patch, the shift in chromaticity of the illuminant, and the resulting final chromaticity of the patch. Little advantage was seen in using a uniform colour space to represent the illuminants in order to more nearly equate the perceptual magnitude of positive and negative values of Δx: making the scale along which the illuminant varied perceptually uniform would not necessarily confer such uniformity on the corresponding variations in the chromaticity of any given Munsell paper. In addition, for most stimuli, the displacements in CIE x, y space were small; the local uniformity of CIE space makes it unlikely that small displacements in one direction would be more perceptible than equal displacements in the opposite direction.

Experiment I

This experiment was concerned with subjects' ability to correctly attribute the changes they saw in the stimulus to material changes or illuminant changes. The subject read instructions which pointed out that the colour of daylight varies from time to time. The instructions explained the experimental task as follows. On each trial, a picture of a pattern of coloured papers would be briefly presented on the monitor and then replaced by a second pattern. The second pattern would either be made of exactly the same pieces of paper as the first pattern, but illuminated by a different phase of daylight, or be made of slightly different colours of paper, but illuminated by the same phase of daylight (the spatial properties of the pattern would remain unaltered). The subject was to press one of two buttons depending on whether they judged an illuminant change or a material change to have occurred.

On each trial, the subject pressed a button to trigger a stimulus sequence. After the sequence had been presented, the subject pressed one of the two response buttons to indicate his response.

The method of constant stimuli was used, with values of Δx in the range $-0.06, -0.05, \ldots, 0.06$ (excluding $\Delta x = 0$). On each trial an illuminant change or a material change occurred with equal probability. Stimuli were administered in blocks of 72 trials, with each stimulus value Δx being represented exactly 6 times in a block, in random order. Three blocks of trials were run in succession with short intervening breaks. Normally one batch of three blocks was performed in a session, lasting about 45 min.

Values of the discrimination index d' (Swets, Tanner & Birdsall, 1961) were calculated for each value of Δx as for a standard "yes–no" experiment. In this analysis, the two kinds of stimulus (material-change and illuminant-change) were considered as generating within the subject internal values drawn from two distributions. The value of d' expresses the separation of these distributions in units of their standard deviation. The advantage of this procedure is that the performance measure d' is not affected by bias on the part of the subjects (a general preference towards making one response rather than the other). The value of d' increases monotonically with the discriminability of the stimuli; $d' = 0$ corresponds to chance performance, and, for comparison, in a 2-alternative forced-choice task, values of d' of 1.0, 2.0 and 3.0 correspond to 76%, 92% and 98% correct respectively.

Experiment 2

The second experiment was concerned with subjects' thresholds for simple detection (as opposed to discrimination) of changes in the stimuli. The experiment differed from Experiment 1 in two ways. First, the range of Δx values was reduced (because the level of performance was higher), and an additional Δx value of zero (i.e. no change) was included. Each value of Δx was represented 6 times in each block of 84 trials. Second, subjects were instructed to press one of two buttons according to whether they thought that a change had or had not occurred in the stimulus during its presentation, regardless of the nature of that change.

Results of Experiments I and 2

Results are presented for the three subjects in Fig. 1. In each case, the discrimination index *d'* is plotted against the magnitude of the shift Δx in CIE *x*-coordinate of the illuminant. The solid symbols represent performance on the stimulus discrimination task (Experiment 1). The open symbols represent performance in the simple detection task (Experiment 2), open circles for uniform illuminant shifts (illuminant-change trials) and open squares for non-uniform shifts (material-change trials). The error bars indicate standard deviations of the estimates of *d'* calculated according to the method of Gourevitch and Galanter (1967). The straight lines are least-squares regressions.

It can be seen that, given sufficiently large values of Δx, subjects performed well above chance in both tasks. The relationship between the levels of performance in the two tasks is considered in the Discussion and in the Appendix. There is no way in which consideration of a single patch in the display could have enabled subjects to perform the task in Experiment l successfully; at the other extreme there was the possibility that subjects may have used some highly global aspect of the display (in which the separate identities of each patch were lost) in order to perform the task. Experiment 3 was performed to investigate this possibility.

Experiment 3

The stimuli used in Experiment 1 were chosen to provide the most straightforward test of subjects' ability to distinguish changes in illuminant from changes of material; the changes observed during the stimulus sequence could most economically be interpreted as being due to one or other kind of change, but not both. One result of this uncomplicated design is that changes in illuminant were accompanied by a shift in what might be called the "space-average colour" of the display, whereas material changes caused no such global change in chromaticity.[5]

It is conceivable, therefore, that subjects may have been able to use the presence or absence of a space-average colour change as a cue for performing the task. Experiment 3 sacrificed some of the straightforwardness of Experiment 1 to determine whether global changes in colour were necessary for the discrimination of illuminant changes from material changes. A change in illuminant occurred on every trial, and a simultaneous change of material occurred only on some trials. Subjects were required to ignore the illuminant change and to indicate those trials on which a change of material had occurred.

The spatial layout and temporal structure of the stimuli were as for Experiment 1. The initial Mondrian was illuminated by a reference illuminant, a daylight having CIE x-coordinate of 0.25 or 0.37 on different trials. Two operations were performed upon the illuminant of the patches to generate the final Mondrian. The first was a uniform shift in the CIE *x*-coordinate of the illuminant on the patches, of magnitude 0.03, 0.06, or 0.09 CIE *x*-units, the shift being positive for the $x = 0.25$ reference illuminant, and negative for the $x = 0.37$ reference illuminant. The second operation was an additional shift of illuminant, in the positive direction for a randomly selected half of the patches and in the negative direction for the remainder. The magnitude of this non-uniform shift (denoted in the figures by Δx) was 0.0, 0.01, 0.02, or 0.03 CIE *x*-units.

As in Experiment 1, appeal is made to parsimony and to the inherent ambiguity of the display to identify the formal operations with physical events involving lights and surfaces. Where the non-uniform shift is zero, the changes observed in the display can be best interpreted as being due to a change in the illuminant; where the non-uniform shift is non-zero, the most economical interpretation is of an additional change in material.

The application of a uniform shift on every trial means that subjects could not use a change in space-average colour to identify illuminant-change trials. Subjects were asked to ignore the changes in illuminant, but to identify those trials on which a change of material had occurred. As in Experiment 1, values of *d'* were calculated for each stimulus condition.

Results of Experiment 3

[5] There may in fact be some change in space-average colour during material-change trials, because the effects of positive and negative illuminant shifts may not cancel exactly. The changes would have been small, however, compared with those occurring during illuminant-change trials.

Figure 2 shows results of Experiment 3 for each of the three subjects. Discrimination index d' is plotted against the magnitude of the non-uniform shift Δx in CIE x-coordinate of the illuminant, for each of the three values of the combined uniform shift in illuminant. The data in the left and right columns of each figure are for trials where the initial illuminant had CIE x-coordinate 0.25 and 0.37 respectively. It can be seen that subjects were capable of performing well above chance at this task. Performance at detecting the non-uniform shift improved as the uniform shift decreased in magnitude.

DISCUSSION

It was proposed in the Introduction that colour constancy might be usefully considered without reference to colour appearance, namely as the ability of a subject to correctly attribute changes in the colour appearance of a scene either to changes in the spectral composition of the illuminant or to changes in the reflecting properties of the materials of which the scene is composed. In Experiment 1, all three subjects showed an ability to make correct discriminations between changes of illuminant and changes of material. Performance increased with increasing shift Δx in the CIE x-coordinate of the illuminant.

For all three subjects performance on the detection task of Experiment 2 was better than that on the discrimination task of Experiment 1, at each value of Δx. Does this difference indicate that subjects are somehow rather bad or inefficient at doing the discrimination task, or is it simply an inevitable consequence of the different decision processes operating in the two tasks? In the Appendix a very simple model is presented of the sort of decision processes that might occur in Experiments 1 and 2. The behaviour of the model suggests that at least part of the differences in d' values for the two experiments (the apparently better performance on the detection tasks) is to be expected simply because of the different demands of the two decision processes.

Note that, in Experiment 1, the value of Δx did not in itself indicate the magnitude of the cue available to the subject in the same way that it did in Experiment 2; the observation that performance in Experiment 1 improved with Δx is not trivially predictable (though such behaviour is exhibited by the model in the Appendix).

The material-change stimulus was chosen in such a way as to be precisely equivalent to the illuminant-change stimulus at a spatially local level. The change in colour of any given patch as the initial pattern was replaced by the final pattern was of the same magnitude in material-change and illuminant-change trials with the same value of Δx. The task could not be performed by making observations on single patches. Only by observing the changes in appearance of several patches conjointly could the discrimination be made successfully. Experiment 3 was performed to ensure that subjects did not simply use the space-averaged colour of the display to perform the task in Experiment 1; the results indicate that subjects' performance in Experiment l was not simply determined by their ability to detect the presence or absence of some overall colour change. A more complex assessment of the relationships between the colour changes undergone by individual patches must have taken place. The nature of Experiment 3 does not allow a corresponding detection experiment to be done, and therefore the efficiency with which subjects performed the task in Experiment 3 cannot be assessed directly. It may be observed, however, that the levels of performance obtained in Experiment 3 are similar to those of Experiment 1, given that the largest nonuniform shift Δx used in Experiment 3 was only half the largest value of the non-uniform (and uniform) shifts Δx used in Experiment 1.

The mechanism or mechanisms by which colour relationships in a scene might be computed for the purpose of colour constancy in humans are not entirely clear at present, but the data presented here suggest that there is a low-level component, which is not based on von Kries (receptoral) adaptation and which is probably pre-attentive: with the relatively brief and successive presentations of the displays subjects did not have the opportunity to perform patch-by-patch comparisons of the initial and final patterns. In a separate 2AFC experiment (Foster, Craven & Sale, 1992) in which illuminant-change and material-change Mondrians were placed on each side of the original Mondrian, subjects were able to perform the task successfully with exposures of less than 200 msec. An appeal can also be made to phenomenology: subjects reported that the decision as to whether a change in illuminant or a change in material had occurred could be made either very quickly or not at all; changes of illuminant tended to be perceived as a coloured wash over the display, whereas changes of material led to a distinctively uneven appearance.

Conventionally, colour constancy has been regarded as a process which, if perfect, would result in complete perceptual invariance of surface colour under changes in illuminant. Perception, however, is not usually an activity performed for its own sake; its main use is to provide information which an organism can use to govern its actions. From that point of view, it may be irrelevant whether the colour appearance of an object does or does not change under a change in illuminant, providing that any changes that do occur can be correctly attributed. Indeed, Jameson and Hurvich (1989) have argued that eliminating the effects of differing illuminants could actually be disadvantageous: the changing colours of a scene could be useful cues to (for example) the weather and the time of day. Some compensation for the colour of the illuminant would be useful, but complete invariance of surface colour may be not only unnecessary, but also possibly undesirable.

The data presented here show that subjects are capable of correctly identifying a scene as being materially unchanged despite a change in the colour of an illuminant, and thus they appear to be largely colour constant according to the approach proposed in the Introduction. The results were obtained in a situation where subjects had no cues as to the nature of the change in the stimulus except those provided by colour changes. In more-everyday circumstances there would be a number of additional cues operating at rather higher levels that would further aid the discrimination of changes in material from changes in illuminant, and thus one might expect even higher levels of competence than obtained here.

In summary, it has been argued that a basic component of colour constancy is the ability of subjects to correctly distinguish changes in illuminant colour from changes in material colour. These experiments have shown that human subjects are capable of making such discriminations highly reliably, and without scrutiny of individual areas of the scene, providing that the change in the illuminant is sufficiently great to be easily detectable.

Acknowledgements

We thank E. R. H. Sale who collaborated in pilot experiments, and J. D. Moreland, M. O. Scase, and S. Westland for critical remarks on the manuscript. We are also grateful to J. P. S. Parkkinen, J. Hallikainen, and T. Jaaskelainen for supplying us with their data for the spectral reflectances of the Munsell papers. This work was supported by an award from the Science and Engineering Research Council Image Interpretation Initiative.

APPENDIX

Theoretical Comparison of Discrimination and Detection Performance

In the Discussion it was noted that performance in simply detecting changes in a Mondrian display was apparently superior to performance in discriminating between different kinds of change, and the question was raised as to whether this difference occurs because of some kind of inefficiency in making the discrimination or because of differences in the kinds of decision processes in the two tasks. In this Appendix, this question is addressed by the development of simple models of the decision processes involved in Experiment 1 (discrimination) and Experiment 2 (simple detection). The relative levels of performance of the two models on the two tasks are compared to those obtained psychophysically in Experiments 1 and 2.

The models are models of signal detection and discrimination and not of colour vision. They are therefore described in abstract terms which could in principle be applied to other perceptual dimensions. The models are intended to give plausible accounts of the two kinds of decision process, sufficient to indicate an answer to the question at hand; they should not be taken as attempts to model exactly the processes involved in human colour constancy.

It is assumed that each patch i of the n patches in the initial Mondrian is assigned a value along a single dimension. These values are drawn from Gaussian distributions with common standard deviation σ.

Changing the illuminant for a particular patch is assumed to cause the patch to be assigned a new value, drawn from a Gaussian distribution with standard deviation σ but with mean shifted by Δ. The difference between the initial and final values assigned to a patch *i* will be denoted by d_i . The mean of the distribution of values of d_i will be Δ, and its standard deviation √2σ.

On illuminant-change trials, the means of the distributions for all patches are shifted by the same amount Δ in the same direction. On material-change trials, half the means are shifted positively and half negatively, by the amount Δ.

The theoretical subject's task on each trial in the discrimination task is to decide which of these two kinds of change (unidirectional or bidirectional) occurred, using only the values of the differences d_i . Traditionally one would want to obtain a maximum-likelihood solution to this problem: to attribute the differences d_i to one or to two distributions depending on which hypothesis predicted the observed values with greatest probability. Unfortunately the extremely large number of different ways in which the values d_i can be assigned to two distributions (for material-change hypotheses) makes computing the associated likelihoods prohibitively lengthy.

Therefore, it was assumed that the subject behaved as if some internal criterion operated. The sign of each change d_i that exceeded the criterion was noted, and all others were disregarded. In the *detection* task, if there was at least one supra-criterion change d_i, the subject reported that a change had occurred. In the *discrimination* task, if all the supra-criterion changes d_i were of the same sign, the subject reported a change in illuminant. If changes of both sign were present, the subject reported a change of material. If no supra-criterion changes occurred, the subject guessed, selecting a response at random with equal probability.

The performance of each model can be observed as a function of the value of Δ (which is analogous to Δ*x* in the psychophysical study). There is one remaining parameter of the model: the value of the criterion; this was set separately for each subject so that the false-alarm rate in the detection task was correct.

The models were implemented as Monte Carlo simulations.

Results

Because the gross behaviour of the models was the same for all three settings of the criterion, we show only one of the three sets of model results. Figure 3 shows the relationship between *d'* values and Δ for the detection and discrimination tasks for a criterion set according to subject BC's performance. The solid curve and the solid symbols indicate the relationship predicted by the models for discrimination performance, and the broken curve and the open symbols indicate the predicted relationship for detection performance. The most important observation to be made concerning these data is that the models confirm the psychophysical finding that values of *d'* for detection are higher than those for discrimination. But the magnitude of the difference is underestimated by the models. (Because of the arbitrary nature of the Δ axis, there need be no conflict between the shape of the theoretical curves and the experimental data of Fig. 1.) A modified version of the models in which the decision was based on a randomly selected subset of the patches yielded larger differences between the two tasks; one might therefore tentatively suggest that the discrimination task is not performed with high efficiency by human subjects (given the colour information available to the decision process). Alternatively, the simplifications of the models, particularly the compression of three-dimensional colour variations into a single axis, are likely to limit the extent to which it can reproduce human behaviour. Nevertheless the models serve as an illustration that the decisions required in discriminating material and illuminant changes are inherently more complex than those involved in merely detecting changes, and as such may require a higher signal level for a given level of performance.

REFERENCES

Arend, L. & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A*, 3, 1743-1751.

Foster, D. H., Craven, B. J. & Sale, E. R. H. (1992). Immediate colour constancy. *Ophthalmic and Physiological Optics*, 12, 157-160.

Gourevitch, V. & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika*, 32, 25-33.

von Helmholtz, H. (1911/1924). *Handbuch der Physiologischen Optik* (3rd edn, Vol. II). Translation by Southall, J. P. C. (Ed.), *Helmholtz's treatise on physiological optics* (pp. 286-287). New York: Dover (1962).

Hering, E. (1920/1964). Collected works in *Outlines of a theory of the light sense* (pp. 7-8). Translation by Hurvich, L. M. & Jameson, D. Cambridge, Mass.: Harvard University Press.

Jameson, D. & Hurvich, L. M. (1989). Essay concerning color constancy. *Annual Review of Psychology*, 40, 1-22.

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.

Krinov, E. L. (1947). *Spectral reflectance properties of natural formations*. Laboratoriia Aerometodov, Akad. Nauk SSSR, Moscow. Translation by Belkov, G. National Research Council of Canada, Ottawa, 1953 (technical translation: IT-439).

Land, E. H. (1977). The Retinex theory of color vision. *Scientific American*, 237, 108-128.

Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America* A, 3, 1673-1683.

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.

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.

Reeves, A., Arend, L. E. & Schirillo, J. (1989). Color constancy in isolated displays. *Perception*, 18, 529-530.

Swets, J. A., Tanner, W. P. Jr & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, 68, 301-340.

Tiplitz Blackwell, K. & Buchsbaum, G. (1988). Quantitative studies of color constancy. *Journal of the Optical Society of America*, A, 5, 1772-1780.

Valberg, A. & Lange-Malecki, B. (1990). "Colour constancy" in Mondrian patterns: A partial cancellation of physical chromaticity shifts by simultaneous contrast. *Vision Research*, 30, 371-380.

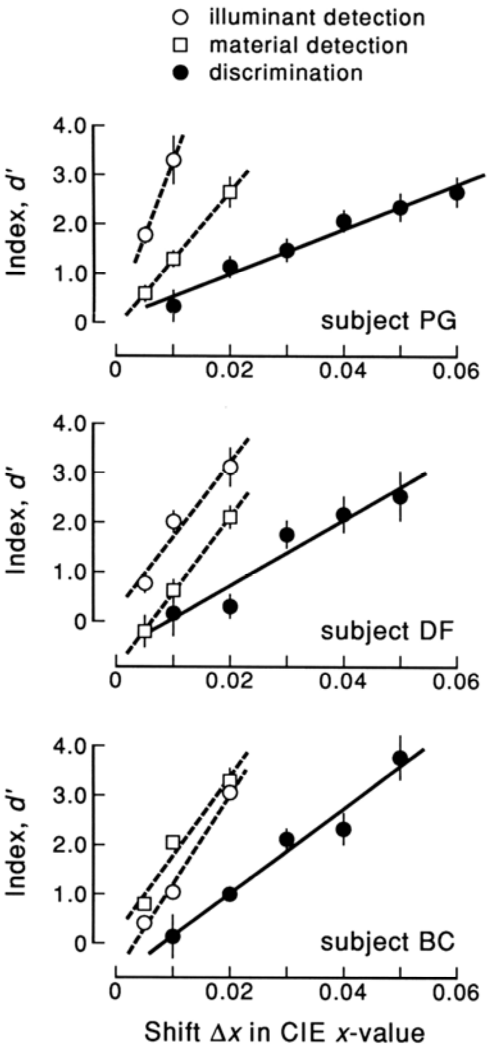

Figure 1. Detection and discrimination of colour Mondrian patterns under changes in illuminant and material as a function of the magnitude of the shift Δx in illuminant CIE x-coordinate (Experiments 1 and 2). Performance in both tasks is quantified by the discrimination index d'. The vertical bars indicate ± 1 SEM and the straight lines are least-squares regressions.

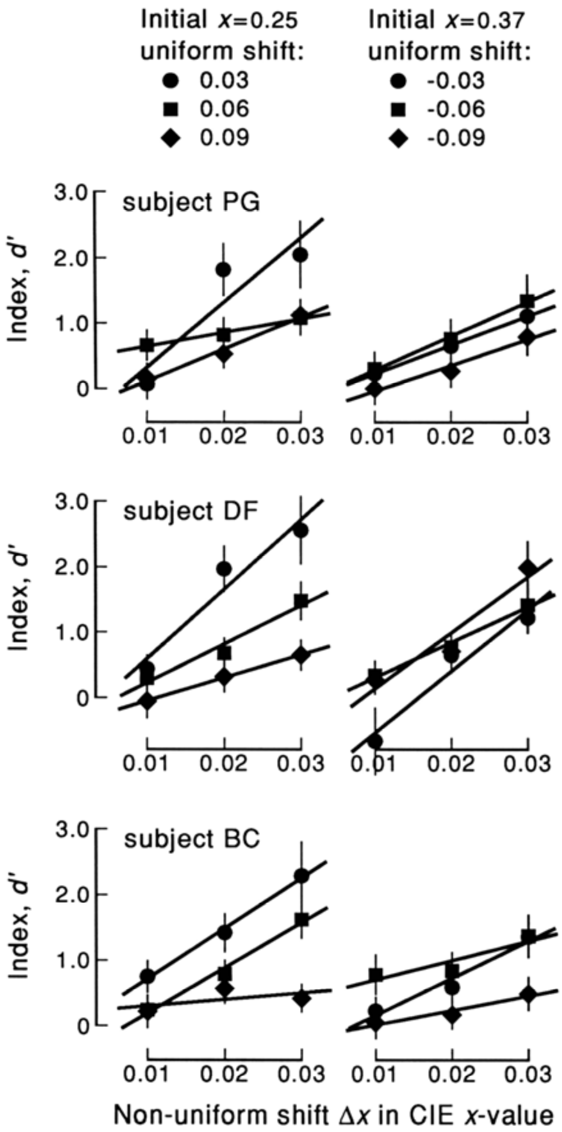

Figure 2. Discrimination *d′* of colour Mondrian patterns under combined uniform and non-uniform illuminant shifts (Experiment 3). Each graph shows performance as a function of the magnitude of the non-uniform shift Δ*x* in illuminant CIE *x*-coordinate, for each of the three values of the combined uniform shift in illuminant, of magnitude 0.03–0.09 CIE units. The data in the left and right hand columns are for initial illuminants with CIE *x*-coordinates 0.25 and 0.37 respectively. The vertical bars indicate ± 1 SEM and the straight lines are least-squares regressions.

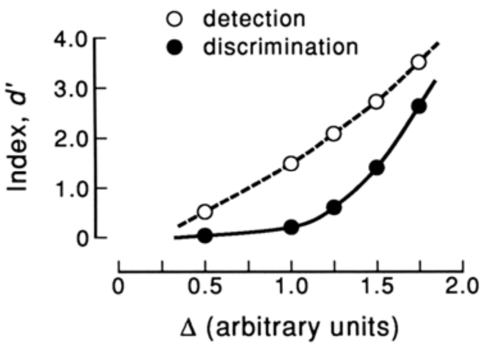

Figure 3. Model predictions of detection and discrimination. The quantity Δ is analogous to the illuminant shift Δx in Experiment 1, but the units of Δx are arbitrary.