The Modulation of Simultaneous Chromatic Contrast Induction

Christopher (Kit) Wolf

Institute of Neuroscience

Newcastle University

Thesis Submitted to Newcastle University

for the Degree of Doctor of Philosophy

October 2011

Acknowledgements

Any PhD is a major undertaking, academically, financially, and sometimes emotionally. Whilst the tasks involved have sometimes seemed overwhelming, it's also true that the time I spent working on my PhD included my most rewarding years at medical school. There are a large number of people without whom it could not have been completed, and to whom I owe a great debt of gratitude.

The Physiological Society, the Barbour foundation and the Ridley family fund provided the financial support for me to undertake my experiments. I admire these institutions for having the vision to support basic science research.

I would like to thank Dr Stephen McHanwell for his support early on in the PhD applications process, and also for his leadership of the BMedSci course, which was stimulating and enthusiastically taught, and was everything I had always hoped University would be.

My inchoate, early experiments were moulded into something much more worthwhile by my supervisor Anya Hurlbert. Marina Bloj was infinitely patient whilst trying to teach me the rudiments of colour theory. Jane Aspell's musical interludes were always welcome, as were Max Hammerton's interesting asides. And it was simply a delight to work with Yazhu Ling. My examiners, Gabriele Jordan and John Mollon made many suggestions for improvements, and it is a far better thesis for their input.

However, I do not mean to downplay the contributions of any of these people when I say that my greatest debt of thanks goes to my friends and family. George and Kerry Holmes made sure I did not get too caught up in my work, and invited me out on walks whenever I was starting to look too pale. Yazhu Ling was both friend and colleague. Pei-Ping Lin made sure that I was eating properly, and more recently Henrike Laehnemann and her musical Saturday afternoons provided a musical accompaniment to the writing-up process. And Kate Akester listened.

My parents realised how important it was to me to complete my thesis and ensured that I was able to take all the time I needed to do so. It goes without saying that one's siblings can always be relied upon to keep one grounded. Helen MacDougall has been a great support to me over the past few years. I step away from the PhD and into the future, with her.

Abstract

A coloured background may induce a contrasting colour in a target figure set against it. This effect, known as 'simultaneous chromatic contrast induction', can strongly influence the way in which we perceive colours and may therefore play an important role in our colour perception. Yet in order to support putative roles in phenomena such as colour-constancy, simultaneous contrast effects must be in some way 'regulated', otherwise they would cause objects to change colour when moved from one setting to another. Here we investigate a number of candidate cues that might be expected to influence the strength of simultaneous chromatic contrast.

In experiment 1, we investigate the influence of global illumination cues that can indicate whether a colour change is consistent with a change in the illumination over the entire scene, or with a localised change in the background colour. Global illumination changes have already been shown to facilitate simultaneous contrast induction, and we further find that this effect is diminished by increasing the luminance contrast between the elements that make up the scene. Contrary to our expectations, increasing the number of distinct chromaticities in the surround does not have any effect. Finally, we attempt to discover whether this effect occurs at a monocular or a binocular site, by presenting different elements of the stimuli either to the same eye as the target, or to different eyes. Here, our results suggest that the effect is probably binocular.

In experiment 2, we investigate how the strength of simultaneous contrast induction changes when textured backgrounds and targets are substituted for the uniformly shaded backgrounds and targets that have been studied in most recent investigations of simultaneous contrast induction. Texture in the background reduces the strength of simultaneous contrast induction for uniform targets. However, texture in targets set against a uniform background strongly inhibits simultaneous contrast induction. When both background and target are textured, simultaneous contrast induction is partially restored suggesting that texture may exert its influence due to segmentation effects. However, another result shows that simultaneous contrast is primarily determined at a local level, arguing against this interpretation of the effect. We do show that the effects of texture are tightly tuned to the cardinal axes, implicating early visual mechanisms in the functioning of simultaneous chromatic contrast effects.

In experiment 3, we investigate two other strong segmentation cues: binocular disparity and differential motion between a target and its immediate surround. Our results show that differential motion may facilitate simultaneous contrast induction, strengthening its effects. However the effect is small and is often inconsistent between observers.

Table of Contents

Table of Figures

Table of Tables

Chapter 1- Introduction

Colour vision enables us to discriminate between lights that differ in their spectral distributions. The physiology underlying the earliest stages of colour sensation is now relatively well understood, yet this alone is not sufficient to explain our perception of colour. For this, we need to understand how colour is represented and transformed at various levels of the visual system.

In this thesis, I concentrate on the role of a phenomenon known as 'simultaneous chromatic contrast induction', by which the perceived colour of an object is influenced by the colours of the surfaces immediately surrounding it. I examine how simultaneous chromatic contrast is regulated and whether it plays a role in other aspects of colour vision, such as colour constancy.

The utility and basic physiology of colour vision

A single photoreceptor type is able to discriminate between two lights only on the basis of brightness, which is to say in terms of the difference in the total amount of excitation elicited by each light. Although the spectral sensitivity of a receptor determines the probability with which it will absorb photons at different wavelengths, once it has absorbed a photon, the physiological response is the same no matter what the photon's wavelength.

For an animal to see in colour, it requires at least two receptor types with differing spectral sensitivities. By comparing their differential excitations when exposed to the same light, it is possible to make discriminations based on spectral composition. In practice, humans with normal vision are trichromats and have three types of cone-photoreceptor. This enables discrimination along two dimensions of colour, and one of lightness.

Most mammals are dichromats and can only distinguish between two dimensions of colour: bright versus dark and blue versus yellow (Jacobs 1993). However, humans are trichromats, possessing a long-wavelength-sensitive cone photoreceptor (L) whose sensitivity peaks at approximately 565 nm, a middle-wavelength-sensitivity cone type (peak approximately 545 nm) (M), and a short-wavelength-sensitive cone type (S) (peak 480 nm). The molecular evidence across species suggests that the L and M cone differentiation arose approximately 30-40 million years ago via a gene duplication event, enabling trichromacy to evolve from the dichromacy of the ancient S cone and a single long-wavelength cone. The genetic coding sequence of the modern L and M cones are 98% homologous (Nathans 1999). The extra class of photoreceptor enables us to discriminate between reds and greens which would otherwise appear identical. However, this apparent improvement in our visual capabilities has hidden costs: an increased sensitivity to chromatic aberration may reduce spatial acuity, and colour-anomalous ("colour-blind") humans, whose visual world is akin to that of dichromats, can sometimes see features camouflaged by red-green patterns that trichromats cannot detect (Morgan, Adam et al. 1992). Nonetheless, trichromacy is highly conserved in those few primate species that have evolved it. Of over 3200 old-world monkeys and apes surveyed, inherited colour-anomalous vision has only ever been found in three closely related individuals (Onishi, Koike et al. 1999; Jacobs and Williams 2001), though on an evolutionary timescale such transmissible deficits are likely to have arisen spontaneously many times over. What tips the evolutionary balance so decisively in favour of trichromats?

Several utilitarian explanations for the evolution of colour vision have been put forward (Mollon 1989). Colour might serve as a cue for object recognition; animals may use colour to assess the health of conspecifics, and colour may aid image segmentation (Rubin and Richards 1982). But the hypothesis that has attracted the most attention is that trichromacy evolved as an aid to frugivory (Allen 1879). This notion is particularly attractive, as many fruits eaten by humans gradually turn yellow, red or orange during ripening. These colours are strikingly visible to trichromats, but dichromats have difficulty distinguishing them from a dappled background of green leaves (Mollon 1989). Furthermore, fruit is an important component of most modern primate diets, and fossil (Kirk and L 2000) and physiological evidence suggests that this was also true of early primates (Milton 1993).

Many tests of the trichromacy-frugivory hypothesis have been based on the observation that all trichromatic primate species investigated so far have remarkably similar photoreceptor spectral sensitivities (Bowmaker, Astell et al. 1991). This is notable because photopigments may undergo rapid evolution (Bridges and Yoshikami 1969; Dulai, Bowmaker et al. 1994), and because they determine the regions of the spectrum to which our colour vision is most sensitive. Ecological investigations have confirmed that trichromacy may make it easier to detect fruit against green foliage (Sumner and Mollon 2000) and to discriminate between ripe and unripe fruit (Sumner and Mollon 2000), but have not shown that primate trichromacy is optimised to see any particular class of object such as fruit, as it is seems well adapted to detecting most things against a leafy background (Sumner and Mollon 2000).

Wolf pointed out that tests of evolutionary hypotheses should take into consideration the fact that different practical tasks make different demands on the visual system (Wolf 2002). The converse is probably also true. Considering the practical utility of colour vision is likely to prove valuable in our efforts to abduce how the visual system is constructed. Three of the most commonly suggested functions of colour vision include segmentation (Zaidi 2000), detection and subsequent identification of objects based on their colour. Colour constancy would seem to be a prerequisite for the latter of these three functions, though not necessarily for the first two.

Colour constancy and chromatic contrast

'Colour constancy' is the tendency of surfaces to appear unchanged in colour despite changes in the illumination spectrum. It is a property only of surfaces in the context of other surfaces. That is, constancy does not occur for a surface viewed 'in the void'. Constancy therefore depends in some way on an interaction between the colour responses elicited in the brain by the different surfaces in the scene (Hurlbert 1998).

The realisation that there was a need for colour constancy arose from investigations of coloured shadows, whose existence implied that light sources may differ in their chromatic composition. Although the phenomenon of coloured shadows was probably known since ancient times (Mollon 2006), their implications were possibly not realised until the late 18th century. Georg Christof Lichtenberg set out the problem in a letter to Goethe, dated 1793 (Joost, Lee et al. 2002), "I take this paper, for example, to be white in the deepest twilight, even at night in the weakest starlight, by tallow, wax or lamp light, in the brightest sunshine, in the red of evening, by snow and rain, in the woods or in a decorated room, etc... we believe at every moment that we sense something which we really only conclude." Hermann von Helmholtz noted that in order to perceive colours as they should be, it is necessary to 'discount the illuminant'. There are different ways in which this may be performed. Smithson (Smithson 2005) describes how Thouless (Thouless 1931) distinguishes 'performance constancy', which refers to an observer's ability to judge the true properties of an object, and 'phenomenal regression to the real object' which refers to an observer's sensations of its properties. Arend and Reeves confirm that it is possible to make different 'paper matches' or 'hue-saturation matches' depending on the instructions provided for the observer (Arend, Reeves et al. 1991). There is therefore a duality to colour perception. We seem able to access both colour-constant and colour-inconstant representations of the world.

Simultaneous chromatic contrast is a phenomenon whereby a surface viewed against a strongly coloured background tends to take on the 'opposite' colour. For example, a gray surface viewed on a red background will appear green, as 'red' and 'green' are complementary colours that lie on opposite ends of an opponent axis – the nature of the opponent axes is something that I discuss in detail later. A more nuanced observation is that simultaneous chromatic contrast acts in such a way that a surface's perceived colour is primarily determined by the cone-contrasts between it and its immediate background, as shown by experimenters such as Shepherd (1997).

This is significant, because under changes in natural illumination (such as daylight) the spatial cone-contrasts between natural surfaces tend to be preserved (Nascimento, Ferreira et al. 2002), and colour shifts where cone-contrasts are preserved tend to be interpreted as illuminant shifts (Foster, Nascimento et al. 1997). Thus, the encoding of colour appearance by cone ratios between surfaces may help to preserve colour appearance under natural illumination changes. A natural interpretation of these findings, and a widely held working hypothesis is that simultaneous contrast is a secondary effect of neural mechanisms devoted to colour constancy (Lucassen and Walraven 1993; Hurlbert and Wolf 2004).

Unfortunately, the corollary is that if simultaneous contrast effects were applied consistently they might cause the colour appearance of an object to change if it is moved from one setting to another, despite there being no change in the spectrum of the illumination and therefore of the light reflected from the object. Despite Zaidi's view that this happens only infrequently (Zaidi 2000), it certainly can and does occur. Yet in general we seem to be unaware of the effect, even though some laboratory demonstrations of simultaneous contrast elicit very large changes in colour appearance.

Perhaps colour contrast is not related to colour constancy at all. Colour may have uses aside from object recognition, such as segmentation (Zaidi 2000), and chromatic contrast that increases the differences in chromatic signals from adjacent areas in the visual field, may conceivably play some role in this. These assertions are by nature difficult to study. Yet another alternative is that simultaneous effects have no aetiological explanation but are somehow a side effect of some other feature of the visual system or a byproduct of its structure. A valid criticism of the stimuli used in vision research is that they are generally not naturalistic. For example, Carandini et al. discuss the failures of models derived from simple stimuli to predict responses to complex stimuli, even early in the visual system (Carandini, Demb et al. 2005). Yuille and Kersten argue that vision works by Bayesian inference; optimised to work on naturalistic stimuli, it should be studied by using stimuli with similar statistics (Yuille and Kersten 2006). It may be that colour contrast, a phenomenon traditionally studied through our responses to simplistic stimuli, has little role in our perception of our everyday visual world in all its rich visual complexity.

A survey of empirical studies of simultaneous contrast

Describing and measuring simultaneous contrast

The first major treatise on simultaneous chromatic contrast was published by Chevreul in 1839. Chevreul was a chemist and philosopher, who worked in the carpet dyeing industry in France in the early 19th century. In the course of his work, he became aware that the colours of his factory's carpets did not appear the way he expected. He realised that the root of the problem was optical or physiological rather than chemical, and after many investigations he wrote his treatise on "The Laws of Harmony and Contrast of Colours" (Chevreul 1868) in which he set out rules such as *"Two adjacent colours, when seen by the eye, will appear as dissimilar as possible"*. His book included illustrations of many of the effects he had noticed, instructions on how they might be constructed and lists of complementary colours. He was almost certainly not the first person to notice such effects: Da Vinci was aware that colours would interact on the canvas (Da Vinci 1954), and Goethe had studied coloured shadows (Goethe 1840). But Chevreul's approach was systematic; his book proved influential, and his theory of optical mixing was later developed by Seurat into the technique of Pointillism (Bleicher 2004).

More advanced research into colour contrast could not simply be based on qualitative observations, but required a means of quantitative measurement. Professor Albert Munsell published a reference book of 'Color notation' (H. Munsell 1905), filled with coloured patches systematically arranged according to hue, lightness and chroma (saturation). With the development of the CIE 1931 colorimetric system, it became possible to numerically express chromaticity, and modern colour research was born (Anon 2007).

Jameson and Hurvich were among the first to undertake quantitative studies into simultaneous chromatic contrast induction (Jameson and Hurvich 1960). They displayed 'test' patches adjacent to surrounds, whose purpose was to induce a perceived chromatic shift in the surface of the test patch, and then they asked the observers to identify a Munsell chip that most closely matched the colour they perceived the test patch to be.

In the same paper, Jameson and Hurvich undertook a further experiment, whose purpose was to determine whether the geometric arrangement of the chips had any influence over the strength of simultaneous contrast induction. They showed that the influence of the surround was strongest if it shared a border with the test patch, and weakened as the surround was moved further from the test patch, or if it was nearly equidistant, but did not share a border. They also compared results across the red-green and blue-yellow subsystems and noted that they appeared qualitatively similar in these respects.

Jameson and Hurvich realised that there were significant drawbacks to the methods they were using. The observer had to glance from the stimulus to the Munsell papers in order to make a match. But more seriously, they realised that this system made it impossible to separate simultaneous and successive contrast (i.e. temporal adaptation) effects. Although real or simulated Munsell papers are often still used in colour experiments, a number of other methods have been developed by which simultaneous contrast may be measured.

Arne Valberg set out to investigate induction and its dependence on the colour attributes of the surround, such as luminance, purity and hue (Valberg 1974). He used a different technique of haploscopic matching, in which a reference stimulus was presented to one eye, and an adjustable stimulus was presented to the other. The observer was able to adjust the colour of the match stimulus until both the test patch and the adjustable patch appeared identical, when viewed against their respective backgrounds. Of course, no method of measuring simultaneous contrast effects is without its drawbacks, and these experiments are dependent on the assumption that contrast works at an early, perhaps receptoral level. Valberg states that the colours of the surrounds, shown to each eye, do not interact with each other – and Chichilnisky and Wandell agree that this effect is minimal (Chichilnisky and Wandell 1995) – although it is difficult to measure it. Some observers also have difficulties with binocular rivalry that can cause parts of the stimulus to disappear and reappear disconcertingly. This is particularly likely to occur during protracted viewing of the stimuli.

Valberg determined that in the photopic range, induction was equally strong when the ratio of the luminance of the central patch to the surround luminance was between 0.3 and 1. The strength of chromatic induction effect increased with time, reaching asymptote after a period of around two minutes. Increasing the width of the surround also increases the strength of induction – but only until the surround diameter reaches 6 degrees of visual angle along each side.

Ware and Cowan (Ware and Cowan 1982) measured the effect of a number of different backgrounds, using a haploscopic apparatus. They then tried to fit a number of different models of receptoral adaptation effects or linear opponent models to the data, to see which provided the best fit. They point out that it is difficult to classify most experiments according to whether they study adaptation or simultaneous contrast, and this is perhaps particularly true for their experiments where it might take as much as two minutes to make each match. They argue that the two processes could well occur in the same neural mechanisms, so they investigated models which were thought to play a role in adaptation, as well as those assumed to play a role in simultaneous contrast. Alas, their results were equivocal and no model appeared to provide a particularly close fit to the data.

One drawback of techniques used to measure contrast is their reliance on short-term colour memory and the often-variable degrees of adaptation that might be evoked when an observer views stimuli under uncontrolled conditions. Krauskopf, Zaidi and Mandler developed a new technique intended to measure simultaneous induction with a minimal influence from chromatic adaptation (Krauskopf, Zaidi et al. 1986). Further, their experimental design avoided the sequential comparisons required by all previous techniques to date. This is desirable as it moves the design closer towards Brindley's 'Class A' category of psychophysical stimuli. Class A stimuli avoid making demands on memory and are therefore thought to be a more reliable measure of sensory processes (Gescheider 1997). The experiment is still not a 'prototypical' class-A experiment as it takes at least a single cycle (two seconds) in order to perform a match. However, due to the constantly-changing nature of the stimuli, any long-term DC adaptation should be to the mean chromaticity of the stimuli, and longer-term memory effects are minimised.

In their technique, the surround colour was continuously modulated along an arbitrary axis, in a sinusoidal fashion. If the target surface were to be left unmodulated, the observer would see its appearance follow that of the surround, but 180^{o} out of phase. This colour shift, induced by simultaneous chromatic contrast induction, may be nulled by an opposing modulation of the target. The observer's task is to adjust a constant that sets the magnitude of the modulation of the target, relative to the modulation of the surround. At the nulling point, the appearance of the target should remain steady, and the observer may proceed to the next match. The size of the constant linking the modulation of the surround and the target patch may be taken as a measure of the size of the induction effect, under a given set of conditions.

Krauskopf, Zaidi and Mandler used this technique for a series of experiments, in which they postulated an early cortical locus for the mechanism underlying simultaneous chromatic contrast, based on data that showed that induction did not take place in independent opponent mechanisms (Krauskopf, Zaidi et al. 1986).

Barnes and Shevell examined specifically the contribution of S-cone excitations to simultaenous spatial contrast (Barnes and Shevell 2002). In this haplosopic viewing paradigm, observers adjusted the appearance of a small comparison patch (0.5^{o}) within a small chromatic surround (1.5^{o}), presented to one eye, to match a 0.5^{o} test patch presented alone against a dark background to the other eye. They noted a lack of interaction between L-M channels and the S-channels, but postulated a cortical mechanism because their findings could only be explained by a mechanism involving both S-on and S-off cells, and few S-off cells were thought to exist at early levels of the visual system. However in a short addendum they qualify this conclusion in light of recent neurophysiological evidence.

Spatial relations influencing contrast induction

Depth and lightness contrast

Although my experiments are principally concerned with simultaneous chromatic contrast, there is a much larger body of literature addressing the problem of simultaneous lightness contrast. Given the obvious parallels between the phenomena, several of these studies are discussed here.

In 1948, Wallach showed that the appearance of a disk, viewed against a background, is determined by the ratio between the luminance of the background and the luminance of the disk (Wallach 1948). Hochberg and Beck investigated whether perceived directions of illumination and surface orientation could influence brightness judgements, and concluded that they could (Hochberg and Beck 1954). Brightness perception is influenced by factors other than the simple ratios between the brightnesses of different surfaces.

Gogel and Mershon (Gogel and Mershon 1969) investigated whether changing the distance between a disk and its circular inducing background influenced the strength of simultaneous luminance contrast. In their experiment, a smaller disk was displayed in front of a larger disk, such that the configuration had a 'center-surround' appearance. The distance between the disks was varied, and observers were able to estimate the displacement based on the binocular disparity between the two disks.

Based on an 'adjacency principle', Gogel and Mershon predicted that the influence of the small disk would be strongest when it appeared close to the larger disk, and this was exactly what they observed. When the disks appeared to be in the same depth plane, the influence of the smaller disk on the colour of the larger disk was almost twice as great as it was when they were separated in depth by 87cm. However, the largest effect of the small disk was only to change the brightness of the large disk by around 25%, so the absolute magnitude of the influence of disparity was relatively small.

Mershon (1971) compared the effects of directional (lateral) and depth displacement, on simultaneous lightness contrast (Mershon 1971). His stimuli differed from those of his previous experiment, in that there was always a degree of lateral separation between the two surfaces, and the level of contrast between the inducing and test surfaces was lower – either 10 or 20 times brighter depending on the condition that was being tested. Both types of displacement weakened simultaneous contrast, directional displacement more strongly than depth displacement. However, the magnitude of the effect remained relatively weak.

Gibbs and Lawson (1973) investigated the effect of stereo disparity on simultaneous contrast, using stereograms, and obtained a negative result (Gibbs and Lawson 1973). They suggested that the small effect observed by Gogel and Mershon (1969) might be due to perceived changes in the relative sizes of the foreground and background stimuli when they were moved closer or further from the observer, or that their findings may apply only to the whiteness illusion (or Gelb effect) but not to simultaneous contrast. This latter explanation could certainly apply to Mershon's 1971 paper. But though Gogel and Mershon claimed to be investigating the Gelb effect in their 1969 paper, in many respects their stimulus resembled that of Gibbs and Lawson very closely and it is difficult to justify the supposition that both groups were investigating different effects, even though they called them by different names.

There were several other differences between the experiments of Gibbs and Lawson, and Gogel and Mershon, that might possibly have accounted for the differences in their results. Gibbs and Lawson collected copious data from only seven highly trained observers, whereas Gogel and Mershon performed only three repeats of their experiments, but tested a much larger number of observers. This may be significant, as Gogel and Mershon observed in their 1969 experiment that the effect of disparity decreased over time. Lastly, there is a small but distinct kink in the graph (figure 5) in Gibbs and Lawson's paper. Although they claim that it is not significant at the 95% level, it appears roughly comparable in size to some of the effects that Mershon did consider to be significant. It is possible that their study simply did not have enough power to confirm a small but extant effect.

In 1977, Alan Gilchrist performed a much more compelling experiment (Gilchrist 1977). He provided two alternative illumination frameworks, and by a minor modification to his test surface he could make it appear to belong to either the darker or the brighter framework. The modification in question was a small nick cut in the edge of a card, which could cause it to appear behind rather than in front of a frame dividing the two illumination frameworks. This minor change to the test surface caused it to be judged as either near-white if viewed in the dark illumination framework, or near-black if viewed in the bright illumination framework.

Was Gilchrist's provision of an additional illumination framework the sole reason that the shift in brightness he demonstrated was so large? One difference between his experiment and those of Gogel and Mershon, was that the contrasts between the stimulus elements in his experiment were larger by an order of magnitude. In Gogel and Mershon's 1969 experiment, there was a 200-fold difference in brightness between the two disks. The lightest and dimmest parts of Gilchrist's stimuli differed in brightness 2000-fold.

The contrast levels that can be produced in dioramas or that are found in everyday scenes are far larger than those typically used in computer-based experiments, which are rapidly becoming ubiquitous (Hurlbert 1998). The contrast ratio of a CRT monitor is limited to around 1000 in ideal conditions, but in general it is much lower. Further, when comparing the luminance and chromatic axes, it is interesting to note that very high contrasts cannot be achieved on the L-M axis due to the overlap in the L and M cone spectral sensitivities. In theory it might be possible to produce similarly high contrasts along the (L+M)-S axis, but in practice most surfaces have relatively wideband reflectances and it would be unusual to find a surface with a sufficiently saturated colour appearance. This implies that even if chroma and luminance mechanisms are similar, phenomena demonstrable using high luminance contrasts may not have chromatic counterparts.

Depth and chromatic induction

Shevell and Miller prepared a stimulus consisting of a ring-shaped test field superimposed against a 2.5 degree wide adapting field (Shevell and Miller 1996). The stimuli were generated on a CRT screen, and were viewed through a stereoscope, and in their basic form consisted of a narrow test ring set against a green local background (they also investigated a more complex arrangement, which I do not discuss further). To measure simultaneous contrast, the observer was asked to adjust the appearance of the test ring until it appeared neither reddish nor greenish, for a range of surround brightness settings. The colours used were those of the CRT phosphors; no attempt was made to isolate induction along a physiological cardinal axis. For each experiment, three basic conditions were investigated with the test ring appearing behind, in front of, or in the same plane as its immediate background. The results showed a small but statistically significant effect of disparity between the immediate surround and the test field. Placing a test surface on a green background should make it appear more red, reducing the amount of 'red' phosphor required to make it appear neutral. Shevell and Miller found that relatively more red phosphor was required when depth disparity was introduced, implying that contrast was weakened in this situation.

Amano, Foster and Nascimento developed a novel means of displaying a test field in a different depth plane from a Mondrian image displayed on a CRT screen, that served as the surround (Amano, Foster et al. 2002). They attached a short Perspex rod perpendicular to the screen, which functioned as a light pipe. By displaying a colour on the screen behind the light pipe, they could cause the end of the Perspex to change colour, and the whole screen and rod were free-viewed, without the need for any form of stereoscope, and bypassing problems related to spatial inhomogeneity of the screen that are a particular problem when using stereoscopes in conjunction with CRT displays.

In their experiment, Amano et al. simulated rapid illumination shifts in the Mondrian, whilst simultaneously shifting the colour of the Perspex surface in ways that were either consistent or inconsistent with a surface reflectance change. The observers' performance at this task was unrelated to the depth between the Perspex surface and the screen, suggesting that depth has very little influence in this task which is closely related to simultaneous contrast.

In short, investigations into the effect of disparity on simultaneous luminance contrast, simultaneous chromatic contrast, and related effects, have shown mixed results. In the majority of cases, changing the depth between the test and surround surfaces had some small effect, but even when statistically significant the practical significance and true causes of these effects must be called into question. However, Gilchrist's paper shows that the perceived spatial arrangement of different surfaces may act very strongly on their perceived lightnesses, at least for achromatic stimuli (Gilchrist 1977).

What could account for this discrepancy? Hochberg and Beck discussed the possibility that there could be at least two different types of brightness constancy, one working at a sensory level and the other at a perceptual level (Hochberg and Beck 1954). They rejected this possibility as being unparsimonious and unprovable, as the distinction between sensation and perception is arbitrary.

Their dismissal seems to me premature: as discussed in the following chapter, there are few or no back-projections from higher levels of the visual system to the retina, so retinal function appears to be relatively isolated from processing in other parts of the visual system. Processes important to colour and brightness constancy such as receptoral adaptation as described by Von Kries, are known to take place in the retina. Yet it seems unlikely that retinal circuitry might resolve the spatial relations that drive our perception of Gilchrist's and Hochberg & Beck's stimuli. And over the past two decades, the notion that early sensory functions might be 'modularised' has become more widely accepted (Fodor 1983).

Gilchrist makes what I consider the most compelling argument as to why he found such a large effect in comparison with previous researchers. He argues that his stimuli are much richer, because they contain at least two independent illumination frameworks, whereas most of the other experiments contain only one. In a scene with only one illumination framework that can place a test surface in context, making that framework appear a little less reliable is unlikely to have any major effect. When two very different frameworks are available, causing one to appear unreliable pushes us to use the alternative framework as our reference. The cue used in his experiment – one based on superposition and contour-completion – is unlikely to be available at a retinal level for reasons discussed later.

Yet Schirillo and Shevell (1993), who provided two alternative illumination frameworks in the form of a light and dark Mondrian in their 1993 experiment, found only a 17% change in brightness settings when they made the test patches appear in the same depth plane as a non-contiguous Mondrian. They attribute this discrepancy to the lack of retinally adjacent coplanar edges within their display.

Discriminating local versus remote effects

While it is clear that the local cone-contrast between a surface and its immediate background has a strong influence on colour appearance, it is not certain to what extent the contrast from more distant surfaces may contribute. The distinction is often made between 'local' and 'global' contrast (Kraft and Brainard 1997), but it is perhaps more accurate to distinguish between surfaces which share edges with the target surface ('local' contrast) and those which do not ('remote' contrast). Where this distinction is made, the contribution of remote contrast to colour appearance has been quantified as much smaller than local contrast (for example, remote fields up to 10^{o} distant from a target surface contribute less than 10% of the induction from its immediate background (Wachtler, Albright et al. 2001; Wolf and Hurlbert 2003). But Kraft and Brainard argue that remote surfaces contribute significantly to colour constancy, even when they contradict the colour specified by immediate local contrast (Kraft and Brainard 1997). Furthermore, Brown and MacLeod show that not only the mean chromatic contrast, but also its variance, contributes to colour appearance (Brown and MacLeod 1997), although their experiment used an adjustment technique and therefore may be studying adaptation rather than short-term contrast effects.

Jenness and Shevell performed experiments where they placed white dots randomly on the background, before asking observers to adjust a coloured ring until it appeared yellow ('unique-yellow') (Jenness and Shevell 1995). They argued that this 'remote context' influenced the yellow setting indirectly, by increasing local induction. Brenner and Cornelissen later disputed this interpretation, showing that if care was taken to retain local cone contrasts between background and test ring, the backgrounds exert the same effect (Brenner and Cornelissen 1998). They also suggest that this experiment is likely to be studying receptoral adaptation rather than simultaneous contrast, pointing out that it takes some considerable amount of time for the effect to become evident.

Wachtler, Albright and Sejnowski made somewhat similar claims relating to the indirect influence of remote squares, but employed a sequential 2-alternative forced choice paradigm rather than an adjustment technique (Wachtler, Albright et al. 2001). This means that chromatic contrast induction is likely to have dominated over adaptation in this situation. They showed that when there was a strongly coloured background present, changing the colours of the remote inducers influenced the colour appearance of a test patch more than it did when the local background was neutral. In other words, the remote surfaces exerted an influence only in the presence of significant chromatic induction. As I replicate this experiment in my first experimental chapter, it is discussed in more detail there.

These findings – in particular those of Wachtler et al. – suggest that contrast induction is liable to be a complex process, perhaps neither amenable to reductionist analysis nor susceptible to influence by higher cortical processes. These are issues that I discuss further in the next chapter.

Spatial and temporal factors influencing colour appearance – distinguishing simultaneous contrast from adaptation

For a light source or coloured patch viewed without context, colour appearance is determined solely by the relative excitations of the three cone types. But in all other situations, it also depends on the context in which that light or patch is viewed. Two phenomena that are fundamental to colour appearance are temporal adaptation and spatial contrast. Temporal adaptation is the alteration in visual sensitivity to a particular stimulus that results from prolonged exposure to that stimulus or a similar stimulus. Chromatic adaptation is usually used to refer to the reduction in sensitivity of specific cone types after prolonged stimulation of the same cone types. However it may also refer to the change in colour appearance of a stimulus following exposure to coloured background (Rinner and Gegenfurtner 2000). Spatial contrast is the change in appearance of a stimulus induced by a spatially surrounding or adjacent stimulus.

Ideally, spatial contrast is an instantaneous phenomenon that depends only on spatial interactions between image regions observed within one fixation. Temporal adaptation requires time for nearby and distant regions to influence each other, as neural response gains are adjusted according to signals sampled over several eye movements and via more slowly propagating lateral interactions.

Empirically, it is difficult to tease apart the two phenomena: the colour appearance of a small target against a large chromatic background, viewed for seconds or longer, will be influenced both by temporal adaptation to the mean chromaticity and luminance of the dominant background and by the instantaneous spatial cone-contrast at the edges between them. According to Rinner and Gegenfurtner, instantaneous spatial contrast effects may contribute up to 60% of the chromatic induction effects of a large uniform background, and do so within the first 25 ms of the background's initial display (Rinner and Gegenfurtner 2000). The remaining induction effects are carried out by two temporal adaptational mechanisms, one fast (with a half-life 40–70 ms), the other slow (half-life 20–30 s). (See below for further reference to these results).

Temporal adaptation

Von Kries (von Kries 1902) proposed that adaptation to coloured lights occurs via a change in receptor sensitivity proportional to its stimulation. In other words, the sensitivity of a receptor is normalised by a scaling factor inversely proportional to its mean excitation. Thus, von Kries explained why a white light will appear bluish to an eye that has been exposed for a long duration to a yellow light.

Von Kries' proposal has been generally translated into the notion that changes in colour appearance following long exposure to non-white lights or surfaces are caused specifically by changes in the sensitivity of individual cone types, rather than by post-receptoral neural mechanisms (e.g. colour-opponent channels or cortical mechanisms). Chichilnisky and Wandell tested this notion using a dichoptic asymmetric colour matching technique, in which each eye was separately adapted to a different uniform chromatic background, and observers matched the colour appearance of a small target presented on one eye's background to that of a target on the other's (Chichilnisky and Wandell 1995). The results were best explained by a model in which the colour appearance of the target is determined solely by the incremental receptor activations in each receptor channel, scaled by a gain factor for each receptor which depends only on the response of that receptor to the background. The incremental receptor activation is the difference between the receptor responses to the target and background. The gain control for each receptor type was best predicted by a simple Weber-Fechner model, in which the gain factor is proportional to the inverse of the background cone excitation plus a constant [i.e. $g_L = 1/(k_L + w_L b_L)$, where g_L is the gain of the L cone, k_L and w_L are constants and b_L is the L cone excitation elicited by the background]. In particular, Chichilnisky and Wandell conclude that the data are best fit by a model that assumes the gain factors for each cone class are independent of excitation of the other cone classes. The observers viewed the backgrounds for at least 2 minutes before making matches. Thus these results support the notion that relatively long-term chromatic adaptation causes a simple sensitivity change in each cone class, inversely proportional to its mean excitation, as von Kries argued. Yet the dependence of colour appearance of the target on the incremental different between its cone excitation and that of the background, is consistent with it being defined by local spatial contrast.

Other studies, for example Rinner and Gegenfurtner (Rinner and Gegenfurtner 2000), examine the changes in colour appearance and discrimination that occur in the early stages of chromatic adaptation to a coloured background, on millisecond and second time scales rather than minutes. They report 'instantaneous adaptation' effects which shift the colour appearance of a small target in the opposite direction to the colour change of the background. Because these near-instantaneous adaptation effects do not affect colour discrimination between targets, they argue that they must not occur retinally but instead must be carried out by cortical mechanisms, and that these are spatial contrast rather than temporal adaptation mechanisms. The two fast adaptational mechanisms (one with a half-life of 40–70 ms, the other with a half-life of 20–30 s) affect both colour appearance and discrimination, suggesting that they are sited most probably in the retina. The time scale and spatial characteristics of the former are consistent with fast, local receptoral adaptation mediated by micro-saccades (Shady and MacLeod 2002). Because the changes in background colour here were always along a cardinal colour-opponent axis (either L-M or S-(L+M)), and because the observers were allowed to make achromatic adjustments only along the same axis, the question of whether adaptation occurs independently in receptoral and post-receptoral channels may not be answered from these results, although they are consistent with independence of adaptation between the post-receptoral channels tested.

Temporal adaptation occurs not only in response to steady backgrounds, but also to variations in the background around a fixed mean chromaticity and luminance – contrast adaptation. Webster and Mollon (1991) found evidence for selective adaptational changes in multiple post-receptoral channels by measuring colour appearance of suprathreshold colours following adaptation to contrast along multiple distinct axes in cardinal colour space.

Spatial chromatic contrast

Spatial chromatic contrast occurs when a surface of one colour induces its opponent colour in an adjacent surface. Typically, the inducing surface is large and entirely surrounds the target surface: for example, a small grey disk acquires a pinkish tinge when surrounded by a large green annulus. Although it is a labile phenomenon which occurs only under particular conditions, it is nonetheless powerful when it does occur. Under optimal conditions, spatial colour contrast is instantaneous, cannot consciously be 'turned off', and almost entirely determines the colour appearance of a surface. For example, in a simple image, the colour appearance of a small desaturated disk can largely be predicted by the spatial cone-contrast (the within-type spatial ratio of cone excitations) between it and its background (Shepherd 1997; Hurlbert, Bramwell et al. 1998).

Qualitatively, therefore, there seem to be at least three distinct sensory mechanisms that interact to determine colour appearance: chromatic adaptation to the mean, which occurs over relatively large spatial and temporal scales, alters the neutral point, and requires at least 1 min to complete after a transition in mean chromaticity (Rinner and Gegenfurtner 2000; Werner, Sharpe et al. 2000); chromatic adaptation to the variance (Webster 1996) which scales sensitivity to cone-contrasts around the neutral point, and requires probably minutes to complete; and spatial contrast between image regions of distinct cone excitations, which occurs 'instantaneously' (Rinner and Gegenfurtner 2000). These mechanisms further interact with colour filling-in processes, so that colour appearance seems to be determined by contrasts spreading away from edges, across regions (Hurlbert and Poggio 1989; Broerse, Vludsich et al. 1999; Rudd and Arrington 2001).

Chapter 2 - The Organisation and Function of the Visual System

The main purpose of this chapter is to outline our present understanding of how information about colour is processed within the visual system. However, many of my experiments involve other stimulus features such as binocular disparity or motion and it also seems pertinent to ask how these are processed. Might we reasonably expect to find significant interactions between the processing of colour and the processing of motion, or are they handled by independent mechanisms?

Divisions of the visual system

In practice, almost any visual task tends to activate large portions of the brain, including areas involved in the planning of eye movements and memory. However, a minimal definition of the visual system would include the retina, optic nerve, lateral-geniculate nucleus (LGN), the 'subcortical pathways' and the visual areas of the cerebral cortex – of which there are at least thirty-two (Felleman and Van Essen 1991) but of which the most thoroughly studied are the early visual areas V1, V2, V3, V4 and V5.

The photoreceptors

Photoreceptors perform the most fundamental aspect of vision: transduction of light impulses into electrical signals that are then processed in the retinal circuitry and ultimately transmitted to the central nervous system.

As previously mentioned, there are three classes of photoreceptor that are important to colour vision: the 'L', 'M' and 'S' cone photoreceptors. These work inefficiently at low light levels, and under so-called 'scotopic' conditions more sensitive 'rod' photoreceptors dominate retinal sensation. As there is only a single class of rod cells, colour vision is extremely limited under these conditions. Although under intermediate 'mesopic' levels of illumination both the rod and cone systems work together, in most investigations of colour vision we attempt to avoid these conditions, where interactions between the rod and cone systems are complex (Stabell and Stabell 1998).

More recently, a fifth 'photoreceptor' class was discovered. These are neither rod nor cone photoreceptors, but ganglion cells that project to the LGN (Hattar, Liao et al. 2002; Dacey, Liao et al. 2005). These neurons seem to respond to absolute light levels, and are probably involved in setting the circadian rhythm rather than being involved in image-forming vision (Berson, Dunn et al. 2002).

The retina

The retina's primary role is clearly in signal transduction. The basic function of the cone and rod photoreceptors in signal transduction is well established, and in recent years many of the basic neural circuits have come to be understood, at least in outline – for a review see Masland (Masland 2001). But far from being a simple peripheral organ of sensation, the retina is developmentally an outpouching of the cerebral cortex (Burt 1993). This is reflected in its substantial complexity: it contains over fifty histologically distinct cell types (Kolb, Linberg et al. 1992; Dacey 2000; Masland 2001), including wide-field amacrine cells that transmit information laterally over long distances within the retina through action potentials. Dacey reports axonal fields of 6 mm diameter in the macaque, which therefore cover a substantial portion of the visual field – at least 10-15^{o} (Dacey 1989). Another cell-type, horizontal cells, also transmit information laterally within the retina and are thought to be involved in forming the opponent channels involved in colour vision (Dacey 2000).

The axonal fibers in the optic nerve arise almost exclusively from retinal bipolar cells, and transmit action potentials to the lateral geniculate nucleus (Burt 1993). The retina is also the origin of separate visual pathways, presumably optimised for the transmission and processing of different types of information (Callaway 2005). The two most studied pathways are the 'parvocellular' and 'magnocellular' pathways. So named for the sizes of the neurons within them, the cells have different response properties. The larger magnocellular neurons are able to transmit information faster and exhibit transient rather than sustained responses. The smaller parvocellular neurons give sustained responses, and are much more numerous. They are also much more responsive to colour, many of them transmitting signals that are chromatically opponent in nature. The presumptive role of the magnocellular pathways are in motion detection and navigation, and the parvocellular pathways are likely to be involved in form and colour vision.

The subcortical visual system

In many nonmammalian vertebrates, the largest part of the visual system consists of a constellation of nuclei that comprise a homologue of the mammalian 'subcortical visual system'. Although over the course of our evolutionary history, growth in the primate cerebral cortex has caused the subcortical nuclei to decline in size relative to the cortical visual areas, they undoubtedly remain important in controlling many aspects of human visual behaviour. The two nuclei of the subcortical visual system that are known to be important to colour-vision are the lateral-geniculate nucleus and the pulvinar nucleus, which both receive direct retinal inputs from the parvocellular 'P_β' cells that carry the opponent colour responses.

The lateral geniculate nucleus (LGN) used to be described as a simple 'relay station', passing information from the retina up to the cortical visual areas (Kastner, Schneider et al. 2006). Recent views are more nuanced, taking note of the fact that a higher proportion of the inputs to the LGN are from the cortex and other subcortical nuclei, rather than the retinae which contribute fewer than 20% of the total proportion of its inputs (Sherman and Koch 1986). The full role of the LGN is still unclear, but several studies indicate that its responses to retinal stimuli can be modulated by cortical activity, specifically attention (Vanduffel, Tootell et al. 2000; O'Connor, Fukui et al. 2002) and binocular rivalry (Wunderlich, Schneider et al. 2005). Yet in contrast to later cortical areas, the LGN maintains a very high degree of organisation – segregating the visual pathways monocularly and maintaining a strict division between parvocellular and magnocellular pathways, which are segregated into distinct laminae (Schneider, Richter et al. 2004). The apparent paradox where binocular rivalry modulates the responses of a monocular neuron is explained by the fact that the modulation is only observed if the cell is already being stimulated – and the stimulating inputs are monocular.

The pulvinar nuclei are much less well studied than the lateral geniculate nucleus. Although Felsten reports finding opponent cells in the inferior pulvinar (Felsten, Benevento et al. 1983), and Cowey, Stoerig and Bannister have labelled what appears to be a parvocellular retinopulvinar projection (Cowey, Stoerig et al. 1994), any retinopulvinar pathway accounts for far fewer of the direct parvocellular colour-pathways from the retina than does the LGN. Following ablation of macaque striate cortex, 80-90% of the P_β cells in the optic nerve degenerate (Cowey, Stoerig et al. 1989; Weller and Kaas 1989); the remaining retinopulvinar pathway may account for the remaining ganglion cells that do not (Stoerig and Cowey 1992; Cowey, Stoerig et al. 1994), though Weller and Kaas question this, pointing out that the proportion of ganglion cells that is lost is age-dependent (Weller and Kaas 1989), whereas the existence of a retinopulvinar pathway presumably is not.

Like the LGN, the pulvinar nuclei have many reciprocal connections both to the cerebral cortex and to other subcortical nuclei. One view is that the pulvinar nuclei mirror the cortex, almost every cortical area being represented somewhere within the nuclei (Shipp 2003). This diversity is reflected in the wide variety of stimulus properties that are represented in the pulvinar, inevitably including colour. However, the most frequently postulated functions of the pulvinar are not related to colour vision; rather it acts as a cortico-cortical relay (Shipp 2003), it may (Bender and Youakim 2001) or may not (Smith, Cotton et al. 2009) regulate visual salience, and it seems to have a role in the coordination of visuomotor behaviour (Grieve, Acuna et al. 2000).

There are many other subcortical nuclei, including large and complex structures such as the ‘superior colliculus’, or ‘tectum’. Although the superior colliculus is important in colour vision for some other animals, such as the rainbow trout (McDonald and Hawryshyn 1999), cells within macaque superior colliculus do not appear to be responsive to colour. Maracco and Li have studied the response properties of 212 superior colliculus cells in the macaque (Marrocco and Li 1977). None appeared to exhibit opponency. 31 cells were also investigated to see whether they exhibited opponent inhibition – none was found to do so. When stimuli were presented against a background with either a similar or a differing colour, only weak indications of simultaneous-contrast like effects were demonstrated. Schiller, Malpeli and Schein have also found little evidence of an opponent input into the superior colliculus (Schiller, Malpeli et al. 1979).

Differences such as these mean that one should exercise caution in making comparisons between the capabilities of the human subcortical visual system and that of other animals. Nonetheless, in many creatures the capabilities of the subcortical visual system are extensive, as can be shown by destroying the structures corresponding to the mammalian cerebral cortex, i.e. the telencephalon (Northcutt and Kaas 1995). Telencephalectomised chicks retain an innate preference for red fruit over red insects, whereas when presented with a choice of a green insect or a synthetic green fruit, they are more likely to peck at the insect (Zachar, Schrott et al. 2008). Experiments discussed later in this chapter show that telencephalectomised goldfish are also capable of exhibiting complex visual behaviour, including making colour-dependent responses to test stimuli (Ingle 1985).

Nuclei involved in light adaptation and the control of pupillary response are also to be found in the subcortical visual system, and the pretectal olivary nucleus (Clarke and Gamlin 1995; Gamlin, Zhang et al. 1995) and Edinger-Westphal nuclei (Burt 1993) in particular are thought to be involved in these processes. If channel-specific regulatory mechanisms were to be found in these areas, they might appear to be a natural locus for colour-constancy mechanisms. The plausibility of this scenario may appear to be supported by the observation that pupils may constrict in response to chromatic changes in stimuli – as well as a wide range of other stimulus features such as motion (Gamlin, Zhang et al. 1998). However, studies made of people with Parinaud's syndrome (Wilhelm, Wilhelm et al. 2002) (damage to the medulla including the pretectum but not the Edinger-Westphal nuclei) and achromatopsia resulting from damage to the cortex (Heywood, Nicholas et al. 1998), both show that the reflex is mediated by corticofugal pathways acting indirectly through the Edinger-Westphal nuclei. Whilst it is difficult to rule out the possibility that there are small nuclei in the subcortical visual system that are important to colour vision, our knowledge of it suggests that other than the LGN and the pulvinar, the subcortical visual system's scope is likely to be restricted to the achromatic domain.

The cerebral cortex

Following the LGN and pulvinar in the visual pathways, is the visual cortex, which appears to be essential to conscious visual perception. For example, there are reports of 'cortical blindness' or 'blindsight' in patients with damage to area V1 (Rees, Kreiman et al. 2002). These subjects claim to have no conscious awareness of any visual perceptions, yet they often retain a startle response and seem able to grasp at objects or to avoid obstacles in their path. Somewhat controversially, more limited damage to other visual cortical areas occasionally seems to cause fairly isolated deficits of colour (Zeki 1990; Bouvier and Engel 2006) or motion perception (Zeki 1991) even though the early stages in colour vision and even colour-constancy remain intact (Hurlbert, Bramwell et al. 1998).

These observations are consistent with the view that there is a high degree of specialisation within the cortex, different cortical areas dealing with different aspects of visual perception. Whilst these conjectures appear broadly to be confirmed by single-cell recordings in the cortex, the results of such studies are rarely unambiguous and are frequently called into question by simple modifications, such as changing the criteria used to classify cells. Visual area V4 presents a classic example of such a controversy. Zeki originally reported that each of 77 cells in V4 was sensitive to colour (Zeki 1973) and put forward the area as a colour-center (Zeki 1980), but Schein and Desimone point out that other investigators obtained figures as low as 20% (Schein and Desimone 1990). They attributed these apparently contradictory findings at least partially to the criteria used to classify cells into 'color' and 'noncolor' classes. Similar controversies exist regarding the chromatic properties of areas V3, some people claiming that a reasonably large proportion of cells are colour-sensitive (Gegenfurtner, Kiper et al. 1997) and others that the area is agnostic towards coloured stimuli (Felleman and Vanessen 1987; Tootell, Nelissen et al. 2004).

Colour is undoubtedly represented in many areas of the visual cortex including V1 and V2, but is there a cortical area specialised for the processing of colour? If so, is it V4? This question is still controversial, but it has become clear that V4 has other roles including form vision and is certainly not solely dedicated to processing colour. Two alternatives that have been proposed, are a dedicated colour-processing area called V8 (Hadjikhani, Liu et al. 1998; Heywood and Cowey 1998), or a more diffuse system of specialised 'globs' spread throughout V4 and neighbouring visual areas (Conway, Moeller et al. 2007). These globs may contain cells dedicated for the processing of colour, even if colour-processing is not the sole or primary function of the particular visual area in which the globs are sited.

Neurophysiological evidence for other aspects of visual specialisation sometimes appears to be clearer. V5 contains a large proportion of neurons that are sensitive to binocular disparity (Maunsell and Vanessen 1983) and which, when stimulated directly, have been shown to influence depth perception (DeAngelis, Cumming et al. 1998). They are also sensitive to motion, but their activity is far more strongly dependent on the magnocellular pathways than on the parvocellular pathways involved in colour perception (Maunsell, Nealey et al. 1990). This is reflected in the fact that relatively few of them appear to be sensitive to colour (Riecansky, Thiele et al. 2005).

Pathways of the visual system

The visual system is commonly outlined in terms of a neat set of processing centres and pathways that form ever more complex representations of the image on the retina as they transmit information through the lateral-geniculate nucleus then up the hierarchy of visual areas in the brain. Though inevitably simplistic, many features of this model seem broadly correct. For example, it is certainly true that the responses of individual neurons become more complex and difficult to characterise as one ascends the hierarchy of visual-processing areas.

Theoretical models of colour constancy often call for information or judgements (Hurlbert 1998) that are extremely unlikely to be formed at early levels of the visual system. For example, colour memory has been purported to play a role in colour constancy, but if a 'banana cell' exists that responds only to bananas of the correct colour (akin to a 'grandmother cell') then it will certainly be situated in the cerebral cortex rather than the retina. Yet, it does not follow that any phenomenon that requires higher cortical functions need only involve mechanisms sited in the higher levels of the cortex. It is possible for information to flow back down the hierarchy, as well as up it. As previously mentioned, although the lateral geniculate nucleus (LGN) is often described as a relay between the retina and the cortex, the majority of its inputs are feedback connections from the cortex rather than feed-forward connections from the retina (Burt 1993). Top-down inputs have also been implicated in creating the 'non-classical receptive fields' found as early as V1 (Zipser, Lamme et al. 1996) and there have been reports that attention is able to mediate the responses of cells in the LGN (McAlonan, Cavanaugh et al. 2008); for a review, see Casagrande et al. (Casagrande, Sáry et al. 2005).

Division into processing pathways

The visual system is also frequently viewed as comprising two main pathways, specialised for different purposes. Early interpretations suggested that there were two distinct processing streams: the ventral stream, predominantly concerned with object recognition and the dorsal stream, concerned with motion perception. In 1992, Goodale and Milner proposed that there are two functionally distinct pathways: the 'what' pathway comprises areas such as V4 that appear to be primarily concerned with object recognition or determining properties of objects, whereas the 'how' pathway comprises areas such as V5 that have a role in guiding actions such as grasping objects (Goodale and Milner 1992). Goodale and Westwood have also reviewed the evidence supporting this position, as of 2004, and they argue that the basic notion of a division of labour into two streams is robust (Goodale and Westwood 2004).

A strict division of the visual system into two completely separate pathways might imply that cues predominantly useful for co-ordinating actions, such as depth or motion, would not be available to areas specialised for the processing of colour or recognition of objects. However, there is little reason to think this is likely to be the case. Firstly, it has never been suggested that the two streams are divided by a firewall; on the contrary there are known to be many pathways linking the ventral and dorsal visual areas. Visual area V4 appears to be particularly prolifically connected, and although it takes the majority of its inputs from parvocellular pathways, Felleman and Van Essen point out that it has major projections from magnocellular-dominated areas such V3, MT, V3A and V4t (Felleman and Van Essen 1991). This has been experimentally confirmed, showing that V4 activity is substantially affected if the magnocellular pathways in the LGN are inactivated (Ferrera, Nealey et al. 1994). Secondly, cells in V4 have been found that are sensitive to binocular disparity (Hinkle and Connor 2001; Watanabe, Tanaka et al. 2002) and although only 13% (Desimone and Schein 1987) to 33% (Ferrera, Rudolph et al. 1994) of cells are usually sensitive to motion, a higher proportion can be induced to develop motion after-effects (Tolias, Keliris et al. 2005). Finally, Felleman and Van Essen point out that their directory of known visual pathways is almost certainly incomplete (Felleman and Van Essen 1991). Given our current incomplete state of knowledge, absence of evidence for a putative connection between two visual sites certainly does not indicate that such a connection does not exist.

'Centrifugal' or 'retinopetal' back-projections

The question of whether there are back-projections to the retina is important to much of my research, as well as several other researchers who have used dichoptic methods such as Chichilnisky and Wandell (Chichilnisky and Wandell 1995) and Steven Shevell (Shevell and Miller 1996). Localising simultaneous contrast effects to the retina or to the central nervous system was a preoccupation of many researchers in the 1960s and 1970s, and many of these also worked on the assumption, made explicitly or implicitly, that central influences on retinal processing were either non-existent or minimal. However, Shevell and Wei entertain the possibility of feedback to the retina (Shevell and Wei 2000). Chichilnisky and Wandell perhaps took the more pragmatic approach of simply performing control experiments to satisfy themselves that there were no functional interactions between visual processing in the two eyes (Chichilnisky and Wandell 1995).

Repérant et al. provide a recent, detailed review of our knowledge of the centrifugal visual system in vertebrates (Réperant, Ward et al. 2006). Such pathways certainly exist in lower vertebrates such as the lamprey, and also in reptiles and birds. There were a number of early optimistic reports of retinopetal neurons in humans, such as that of Wolter and Knoblich in 1965, who studied the pattern of neuronal degeneration in the optic tract of a man who had had a bilateral enucleation fifty years previously (Wolter and Knoblich 1965). They estimated that a full 10% of neurons in the optic-tract were retinopetal, and felt that a number of distinct neuronal populations were present. Their methodology was later criticised by Brindley and Hamasaki, who sectioned the optic-nerve of a cat without enucleation and found that there was no difference in the form of degeneration observed, even though any putative centrifugal neurons would also have been cut (Brindley and Hamasaki 1966). More recently, Usai et al. identified a population of axons in macaque retina that appeared to radiate from the optic disc (Usai, Ratto et al. 1991). They noted that they had large axons, were few in number and had large receptive fields. However, their study was confined to the retina and was insufficient to determine the origin of the pathways. Peterson and Dacey showed that some retinal ganglion cells have axons that enter the optic disk, then loop back to synapse with the retina (Peterson and Dacey 1998). Although morphologically their cells differed from those described by Usai et al., this emphasizes the possibility that the cells that they reported may originate in the retina, until such at time as their origin is definitively traced.

Having shown that retinopetal pathways probably exist, it is natural to consider their putative functions. Unfortunately this is largely a matter of conjecture, even for birds and lamprey where the pathways have been recognised for much longer than in other vertebrates. Additionally, at least 8 distinct retinopetal pathways have been described and they appear to vary greatly between species (Réperant, Ward et al. 2006), suggesting that conclusions based on studies of non-primates should only be made tentatively. At present, the best option appears to be to attempt to guess their likely function based on their hodology. Extrapolating from findings in other mammals, the two most likely origins for the human centrifugal system are the hypothalamus and the pretectal area (Réperant, Ward et al. 2006), which is known to be involved in mediating light adaptation (Burt 1993). Perhaps with this in mind, Usai et al. suggested that the neurons they described would probably be suited to a role in retinal adaptation. It's also important to emphasize the small number of neurons that have been reported. Repérant and Gallego suggested that as few as 10 neurons of this type may be present in each retina (Reperant and Gallego (Réperant and Gallego 1976), cited in Simon et al. (Simon, Martin-Martinelli et al. 2001)) and several studies in primates and rats have found comparable numbers – for discussions see Simon et al. (Simon, Martin-Martinelli et al. 2001) and Repérant et al. (Réperant, Ward et al. 2006). One possibility is that the neurons are vestigial, but if they do have a function then it is almost certainly global in scope rather than as part of a spatially selective mechanism. However, in Aves, the dorsal isthmo-retinal system is thought to be involved in attentional mediation of retinal function. Human studies of the influence of attention on the electroretinogram have been performed and although Oakley and Eason (Eason, Oakley et al. 1983) have reported positive findings, these results remain controversial and other teams such as Mangun, Hansena and Hilyard have failed to replicate them (Mangun, Hansen et al. 1986).

In summary, although evidence for the existence of centrifugal projections to the retina is sparse and is often challenged, it seems premature to conclude that they are either non-existent or non-functional. However, early reports that a large proportion of neurons in the optic nerve were retinopetal have not been substantiated, and their low number suggests that in functional terms any input to the retina seems likely to be spatially non-specific.

Colour-representation in the retinal and central visual pathways

Single-cell recording of retinal, lateral-geniculate-nucleus and cortical responses to chromatic stimuli have provided a wealth of knowledge about how colour and brightness information are coded at different anatomical sites within the visual system.

The earliest studies in this area investigated cell responses in the LGN, with the first study carried out by DeValois, Abramov and Jacobs in 1966 (DeValois, Abramov et al. 1966). It was quickly established that cells here exhibit opponent responses oriented approximately along the cardinal axes, with centres that respond to single cone inputs and surrounds that respond to the opposing inputs. Four chromatic opponent cell types were described: +R-G (red excitatory, green inhibitory), +G-R, +Y-B, +B-Y, as well as two types of non-opponent cells (DeValois 1981). More recent studies have broadly confirmed this view (Derrington, Krauskopf et al. 1984; De Valois, Cottaris et al. 2000), with the main controversy probably being whether the contribution to the antagonistic surround appears to be cone-specific (Reid and Shapley 1992) or whether it is drawn from random contributions of L and M cells – probably a result of the randomness of the retinal cone-matrix (Solomon and Lennie 2007). In either case, the resulting opponent pathways would be oriented along approximately the same cardinal axes.

Single cell recordings from bipolar cells in the retina have now also been made, and their findings broadly mirror those of studies of the LGN in that they already show clearly defined opponent pathways of both ON and OFF types. Although initial reports found only centre-on S-cone pathways, centre-off S-pathways have recently been reported, as in the LGN. There is also a clear histological distinction between the bipolar cells involved in LM, S-opponent and luminance pathways (Dacey and Packer 2003).

In visual area V1, the situation changes abruptly. The cells are no longer specifically tuned to the cardinal axes – if anything the cardinal axes appear to be under-represented. The responses are nonlinear causing a sharpening in the tuning of the response. S-off cells also become as common as S-on cells, which is not the case in either LGN or in the retina (De Valois, Cottaris et al. 2000). One possibility is that cardinal axes are still represented within V1, but by cells that are hidden amongst the wider population of less specific cells. However, even when cells are categorised into simple or complex cells, or by layer, no population of cells that is specifically tuned to the cardinal axes seems to emerge (Lennie, Krauskopf et al. 1990).

Visual area V4 is often implicated in colour constancy, both because it is thought by some to contain a high proportion of colour-sensitive neurons and because they are known to have extremely large receptive fields (Schein and Desimone 1990), the ability to integrate information over large areas of the visual field being a feature that is assumed to be necessary in many models of colour-constancy. A recent study by Stoughton and Conway has found evidence that some neurons in this area respond to the 'unique hues' of red, green, yellow and blue and that they also exhibit the Bezold-Brücke effect, whereby the perceived colour depends in part on its luminance (Stoughton and Conway 2008). Whether or not one accepts that cortical areas V4 or V8 are specialised for colour, the view that their responses to colour are more sophisticated than those of lower visual areas seems inescapable.

Neural correlates of simultaneous contrast

Where in the visual system is simultaneous contrast induction first represented? A natural way of asking this question, is to search for neurons that seem to exhibit simultaneous contrast. Zeki, in 1980, appeared to have found such evidence when he described cells in V4 that appeared to respond in a colour-constant way to Mondrian stimuli (Zeki 1980). Although he does not claim to be investigating simultaneous contrast, his stimuli – presented for short durations – invoke relatively instantaneous spatial comparisons rather than long-term adaptation. As discussed in the previous chapter, Krauskopf, Zaidi and Mandler also argue in favour of a cortical locus on the basis of psychophysical evidence (Krauskopf, Zaidi et al. 1986).

Shevell has argued that perceived depth may modulate the strength of simultaneous induction effects (Shevell and Miller 1996). If so, this does not necessarily disprove the assertion that contrast occurs at a low level of the visual system; very many feedback connections link higher visual areas dealing with shape and lower visual areas such as monocular layers of V1 or the lateral geniculate nucleus. However, in a separate paper, he argues strongly for the cortex, given that an articulated surround appears to influence chromatic induction equally whether it is presented to the same eye, or a different eye, from the target of induction (Shevell and Wei 2000). The task used in both these papers was an adjustment task of uncontrolled duration, hence it is possible that he was studying adaption rather than simultaneous contrast induction.

Cerebral infarction sometimes destroys discrete areas of the brain, destroying some faculties whilst leaving others intact – as is the case in cerebral achromatopsia. There are several reports of patients who have impaired colour-constancy following damage to higher visual cortex. Kennard, Lawdon, Morland and Ruddock report the case of 'BL', a patient with incomplete cerebral achromatopsia (Kennard, Lawden et al. 1995). The fact that he retains a degree of colour-vision means that it is possible to subject him to simple tests of colour-constancy, involving colour-naming for patches displayed as part of a Mondrian display. In contrast to colour-normal observers who always chose the same colour categories whatever the illuminant, BL's choices changed dramatically under different illuminants, and his choices of colour terminology could be predicted quite accurately based on chromaticity alone.

A more recent study performed by Clarke, Walsh, Assal, Schoppig and Cowey examined seven patients with focal neurological damage to the prestriate cortex, and forty-six colour-normal controls (Clarke, Walsh et al. 1998). Their experimental method was based around colour-matching under free-viewing, but they attempted to use more naturalistic stimuli, simulating realistic changes in illumination. Of the seven patients, three appeared to exhibit deficits in colour constancy similar to that of BL, despite the fact that none of them were suffering from any residual achromatopsia at the time of the experiment.

Bramwell, Hurlbert, Heywood and Cowey have also studied an individual 'MS' with cerebral achromatopsia who nonetheless appeared to compute cone-contrast ratios correctly when undergoing a range of psychophysical tests. They worked on the assumption that any breakdown in the computation of cone-ratios should cause changes in the relative brightnesses between surfaces, even if MS could not perceive their colours directly. This could be tested by asking their observer, MS, to perform matches-by-adjustment or odd-one-out tasks. They concluded from his results that his visual system does compute local cone-ratios, and that this is likely to be done early in the visual system, perhaps at a retinal level. However, they also noted that he almost ignored S-cone inputs – in contrast to the patients previously mentioned who were able to perceive S-cone inputs, but who did not have S-cone colour-constancy. He was unable to compute global cone-contrasts, and could not perform tasks based on complex images rather than simple center-surround stimulus configurations. These findings are consistent with the concept that colour-constancy is a multi-stage phenomenon, but should weigh against the view that a double-dissociation between colour-perception and colour-constancy has been found.

Despite the evidence supporting the cortex as the locus of simultaneous contrast, the issue is not uncontroversial. Unlike Shevell et al. (Shevell and Miller 1996; Shevell and Wei 2000), Amano, Foster and Nascimento failed to find any links between depth perception and simultaneous contrast (Amano, Foster et al. 2002). The neural mechanisms underlying simultaneous contrast, as demonstrated in its simplest form, have been shown to be sited at early monocular levels of the visual system. Shepherd claims that a background presented to one eye does not influence the perceived colour of an object viewed by the other (Shepherd 1997); they must be presented to the same eye if the background is to significantly influence the appearance of the test object.

Another way of considering the putative capabilities of the subcortical visual system and retina, is to ask whether 'lower' vertebrates possess colour constancy and whether their sensations are influenced by simultaneous contrast. Jacobs points out that such experiments are notoriously difficult to carry out properly: even the question of whether animals have colour vision was initially difficult to investigate behaviourally (Jacobs 1981). However, Gnybukin et al. suggest that toads possess colour constancy (Gnyubkin, Kondrashev et al. 1975) (discussed in Neumeyer (Neumeyer 1998)) and Intskirveli, Roinishvili and Kezeli have also shown that guppies also possess colour constancy (Intskirveli, Roinishvili et al. 2002). Ingle (Ingle 1985) and Dörr and Neumeyer (Dorr and Neumeyer 2000) have separately demonstrated both that goldfish possess colour constancy and that their perception is subject to chromatic contrast (Ingle 1985; Neumeyer, Dorr et al. 2002). In his paper, Ingle further points out that ablation of the telencephalon, which in mammals has evolved to form the cerebral cortex (Northcutt and Kaas 1995), does not interfere with this function (Ingle 1985). Computational work by Kamermans, Kraaij and Spekreijse suggests that simultaneous contrast emerges when one models the behaviour of well-studied horizontal cells. It therefore appears possible that simultaneous contrast is computed at a retinal level, at least in the goldfish (Kamermans, Kraaij et al. 1998). Kamermans et al. also address the issue of why other researchers may not have been able to find neurons in V1 that were subject to simultaneous contrast. They argue that if their model is correct, groups of neurons will be subject to contrast induction, but individual neurons will not.

The relevance of these observations depends on the degree of similarity between the teleost retina and the primate retina. Joselevitch and Kamermans review our knowledge of the teleost retina, directing particular attention to this question (Joselevitch and Kamermans in press). They argue that although there are many differences between teleost and mammalian retina, they are ultimately similar in their basic architecture.

These studies clearly do not rule out the possibility that human simultaneous contrast is computed cortically. Yet they are consistent with the notion that chromatic contrast might potentially be carried out very early in the visual system. It is exciting to consider that the neural mechanisms of simultaneous contrast might soon be understood in teleosts, and if this is accomplished it will greatly increase the importance of studying whether the primate retina functions in a similar manner.

Binocular recombination and depth representation

Due to the physical phenomenon of binocular disparity, each eye sees a slightly different picture from the other. At early levels of the visual system the representations of the view from each eye remain separate, but at higher levels they are combined to form a single representation. This difference in the visual images due to binocular disparity may be used to obtain information about depth. As depth can only be computed after binocular recombination has occurred, an interaction between depth and other perceptual phenomena would suggest a central locus for their mechanisms.

At least two forms of disparity may be calculated within the visual system. The first is 'absolute disparity', related primarily to the fusion angle required to combine the two visual images. This provides a good indication of absolute distance from the observer. The second cue is 'relative disparity' whereby after an image has been grossly fused, the order in which surfaces are positioned relative to one another can be determined without necessarily determining their absolute distance. As discussed in the methods chapter, in experiments using a stereoscope and a computer monitor it is easier to generate relative disparity than absolute disparity.

Read provides a recent review of early mechanisms of binocular disparity (Read 2005). There appears to be little evidence for any binocularly sensitive neurons in the precortical pathways, but although she points out that many cells in V1 have been found to be sensitive to disparity, disparity and depth perception are not synonymous as neurons in V1 respond to locally randomly correlated stimulus features even when, at a gross level, the stimuli are not correlated. Although their responses are admittedly weaker for such uncorrelated stimuli, there is some debate over whether this is due to feedback from higher areas that do respond only to gross disparity (e.g. Ohzawa 1998) or, as Read suggests, a thresholding mechanism operating without feedback (Read, Parker et al. 2002). V1 cells also respond only to absolute disparity rather than relative disparity. These features, taken together, suggest that binocular disparity is only poorly represented at this early level of the visual system.

Parker provides a more comprehensive overview of depth-perception mechanisms in the visual cortex, including higher visual areas (Parker 2007). Disparity sensitive neurons that respond primarily to true differences in depth are not found until V2, where neurons that respond to depth differences in outline contours have been identified. Their response to uncorrelated disparity is much lower than their response to correlated disparity.

Visual area V4 has already been discussed as a putative site of colour constancy, and the binocular responses of its neurons are therefore of particular interest. Hinkle and Connor have shown that cells in V4 are sensitive to disparity (Hinkle and Connor 2005) and Umeda et al. have shown specifically that they respond to relative disparity (Umeda, Tanabe et al. 2007).

An important caution that should accompany these observations is that feedback pathways may enable representations of depth perception to influence other aspects of processing in lower centers in the same way that binocular disparity is known to influence activity of monocularly driven cells in the LGN (Wunderlich, Schneider et al. 2005). Equally, it is theoretically possible that colour pathways may remain independent of depth representation even if other mechanisms within the cortical areas they pass through are able to decode binocular disparity cues. Although it is clearly still desirable to know as much as possible about how and where depth is represented in the visual system, these caveats mean that any association between the neural locus of depth perception and simultaneous contrast should be made with care.

Motion perception

The stimuli in my experiments invoke the perception of relative motion. That is, there are multiple surfaces within a display, whose directions of motion appear to differ from one another. There was never any intention of causing the observers to perceive absolute motion – that is that the observers themselves were moving relative to the stimulus. Smeets and Brenner provide a good diagram demonstrating the differences between these various types of motion (Smeets and Brenner 1994). In all my experiments, the stimuli exhibited normal relative motion, although in some cases only the larger stimulus (ground) moved whereas in others only the figure moved. Despite Smeets and Brenner's assertion that absolute motion perception is abolished whenever a fixed reference is visible, many of my observers complained of 'seasickness' during the moving-background conditions.

A further distinction is that between first-order and second-order motion. First order motion is defined by high contrasts between a moving object and its background, whereas second-order motion perception is defined by other features such as texture that do not influence the first-order statistics of the stimuli.

There is some evidence that first-order and second-order motion are likely to be processed at least partly independently. Vaina and Cowey report a patient whose responses to second-order motion is impaired, but who can still perceive first order motion as well as control observers (Vaina and Cowey 1996). Demonstrating a double dissociation, they also report a patient with an impairment of first-order perception but not second-order perception (Vaina, Makris et al. 1998). Although early imaging studies were inconclusive, some studies have found evidence that first and second order motion are processed by different mechanisms (Dumoulin, Baker et al. 2003), if not necessarily at different loci (Ashida, Lingnau et al. 2007). Psychophysical and neurophysiological evidence is also consistent with this view. Allard and Faubert point out that first-order masking does not mask second order motion, and second-order masking does not interefere strongly with first-order motion (Allard and Faubert 2007). Single-cell recording appears to suggest that second-order motion and first-order motion are represented by different neurons, first order motion being represented as early as V1 and second-order motion first being represented in V3, but in his review Baker finds no evidence for cells specialised exclusively towards the processing of second-order motion and points out that cells responding to second order motion can be found in V1 (Baker 1999) – something also reported by O'Keefe and Movshon (O'Keefe and Movshon 1998). Another approach comes from the study of visual popout – Ashida, Seiffert and Osaka show that although first-order motion can be detected without extensive visual search, second-order motion requires a consecutive search (Ashida, Seiffert et al. 2001).

A final comment on motion processing, is to reiterate that motion is represented in many extracortical loci, including the superior colliculus, and even to some extent in the retina. Berry, Brivanlou, Jordan and Meister have shown that the rabbit retina can 'anticipate' the motion of a moving bar (Berry, Brivanlou et al. 1999), and Olveczky, Baccus and Meister later suggested that this might be involved in figure-ground segmentation based on differential motion (Olveczky, Baccus et al. 2003).

Conclusions

The main purpose of this chapter was in part to overview the visual system so that I can later relate my findings to known neuroanatomy and physiology. There are four main possibilities for the locus of the simultaneous contrast mechanisms: the retina, the cortex, the subcortical nuclei or some combination of these, acting in concert. Of these options, none can be entirely ruled out based on current evidence. Although colour-constancy is widely assumed to be a complex process and it therefore seems reasonable to expect the mechanisms involved in regulating chromatic contrast-like illusions to be equally complex, it is important not to underestimate the capabilities of the early visual system including the retina.

I also posited the question of whether, given that many cues are thought to be processed in specific, different areas of the brain, it was reasonable to search for interactions between them. The main features of the visual system that emerge over the course of this chapter are its almost overwhelming complexity and interconnectedness. In view of this, if any cue may putatively be used in colour-constancy there should be no a-priori reason to assume that it will not be available to those sites that are particularly important for colour vision. However, these same features mean that any conclusions about the likely neural locus of mechanisms involved in colour-constancy and contrast effects should be made with caution.

Chapter 3 - Methods

Psychophysics

The term ‘psychophysics’ was coined by Gustav Fechner around 1860 (Fechner 1860). The basic philosophy of the field is that it is possible to infer the organisation and function of our sensory and perceptual systems through the investigation of behavioural responses to carefully designed physical stimuli. As such, it is a cheap, elegant and non-invasive way of investigating the visual system.

Over the past century psychophysics methods have been developed considerably, in order to reduce the effects of bias, to increase the efficiency of experimentation, and to quantify the reliability of the methods employed. Gescheider overviews many of these developments (Gescheider 1997). In this thesis I have employed primarily the method-of-adjustment and method of constant stimuli, to be described in due course. Both ultimately probe the same underlying physiological construct: the psychometric function.

The psychometric function

The psychometric function describes an observer’s subjective responses to a varying physical stimulus. Our sensory system is limited in its ability to detect or discriminate any stimulus. What is represented in the psychometric function is that when confronted with an ambiguous stimulus, our decisions on what it shows are probabilistic. As a stimulus becomes stronger, or the difference between two stimuli becomes greater, our ability to detect this improves, as does our subjective confidence in our judgements.

The main parameters of a psychometric function are its threshold, slope, and lapse rates. For my experiments, the threshold is the most important parameter. Thresholds can be reported for any level of performance on a psychometric function. However, as I am predominantly interested in determining ‘points of subjective equality’ (PSE), I always employ the 50% threshold.

The other parameters of the psychometric function are largely ignored. I rarely report the slope, although it is clearly one of the factors related to the confidence-intervals reported for each threshold. The lapse and false alarm rates refer to errors made by an observer, even when stimuli are unambiguous. In some situations there is a qualitative difference between lapses made at different ends of the psychometric function, but for my experiments there is no obvious reason an observer should make one type of error more frequently than another, and they are sometimes tied together to make a single error-rate estimate. They can significantly change estimates of p-values calculated using pfcmp, as is discussed in the following chapter, and if one is willing to make the assumption that both lapse-rates are likely to be identical, 'pfcmp' becomes a significantly more powerful statistical method.

Measuring the psychometric function

All the experiments in this thesis aim to discover a 'point of subjective equality' where a stimulus viewed against an inducing background is judged identical to a neutral reference.

Method of adjustment

In the method of adjustment, an observer is asked to adjust a stimulus along one or more continuous axes until its appearance satisfies a predetermined criterion – perhaps appearing identical to a reference stimulus, or in experiments reported here, until its colour appears to be constant.

From repeated measurements, it is possible to calculate the standard error and standard deviation of the observer's performance at the task. This is closely related to the slope of the psychometric function, but the information can be used to set confidence intervals for the threshold, which is calculated as the mean of the measurements.

Method-of-adjustment experiments are relatively simple to set up, as they do not need to be individualised and are therefore less labour-intensive for the experimenter. However, they do not directly measure the psychometric function, and factors such as the spread of an observer's settings can depend as much on his willingness to accept an imperfect match as on his sensory or perceptual ability.

Method of constants

The method of constant stimuli consists of measuring the observer's performance on a psychophysical task, at a small number of distinct threshold intensities.

In the first, preliminary, phase of experimentation the observer's threshold is approximately localised. In my experiments this was performed using an 'expanded' distribution of constants that ranged considerably higher and lower than the anticipated threshold position. By examining 0% and 100% performance points, it is possible to make an approximate guess as to the location of the point of subjective equality.

The next step is to choose a new selection of constant reference points at which to sample the psychometric function. For my initial experiments, I ensured that the range of settings included two points where performance was likely to be near 0% or 100%, but that were just on the boundaries of the tails of the sigmoid function. I distributed at least four remaining points roughly equally between these extremes, ensuring that at least some of the points were near the 50% threshold for the data. Sometimes it was necessary to repeat this step until the observer's psychometric function was approximately localised.

Intuitively, the simplest psychophysical task to determine a 'point of subjective equality' would be to show an observer two coloured surfaces and ask him to judge whether they were the same colour or different. However, observers differ in their willingness to accept as identical surfaces that always differ slightly. This can cause the spread of the function – which will be Gaussian-like in form rather than a sigmoid curve – to depend more on the observer's assiduousness than on their discriminatory ability. For this reason, such simple 'same/different' tasks have been superseded by forced choice techniques.

Although we sometimes call my experimental methods a ‘2 alternative forced choice’ (2AFC), they differ very significantly from the usual form of these experiments. My experimental methodology is derived from Wachtler et al. (Wachtler, Albright et al. 2001). Classically, two alternative stimuli are presented and the observer is asked to identify one of them as being the ‘target’. For example, in an experiment to determine the smallest perceptible colour difference, two alternatives might be given, along with a reference. The observer is asked to identify which of the alternatives is a different colour from the reference. Although a wayward observer might be able to bias his responses by failing to select stimuli that he recognised to be a different colour, he could not bias the experiment by adjusting his readiness to answer ‘different’ when the colour difference is too small for him to be certain of what he has seen. When this is the case, he will simply perform at chance – correctly selecting the ‘different’ stimulus only half of the time. This limits a consciously or unconsciously biased observer’s ability to influence the results of the experiment.

In my experiments, we presented only a single stimulus, and asked observers to identify whether it appeared one of two colours (e.g. red versus green). Although they are forced to select one of two alternatives, the alternatives pertain to two mutually exclusive attributes of a single stimulus rather than representing a choice between two different stimuli. This is the crucial difference between this method, and a classical 2AFC experiment. The method I use solves the problem of the ‘same/different’ task – that of an observer’s willingness or reluctance to judge two stimuli to be identical. But it does not address the second problem of bias. In my experiment an observer could influence the experimental threshold if, consciously or unconsciously he was biased towards answering choosing one alternative over the other.

Estimating the parameters of the psychometric function

In my first experiment, thresholds were determined by a proprietary program written by Bruce Cumming in 1990. It fits a cumulative Gaussian function to the raw data, using a maximum-likelihood technique based on that described by Watson and Pelli, who described it as part of their 'Quest' adaptive method (Watson 1983). This technique finds the threshold value which maximises the log likelihood function, i.e. the probability density function of the data conditional on the threshold value. Here we used the mode of the cumulative Gaussian function as the threshold value. The 95% confidence interval limits on the threshold value were also calculated by the fitting program according to the technique described in Watson & Pelli. Their technique exploits the fact that a multiple of the log-likelihood function is asymptotically chi-square, and therefore we construct an interval beyond which thresholds may be rejected as sufficiently different from the most likely threshold, at a given level of significance.

However, when I undertook my first series of experiments, I was unable to obtain any documentation on Bruce Cumming's program. Neither did I have access to the source code in order to determine how it fit curves, or estimated their variance. For future experiments, including the second series of experiments in chapter 4, I therefore decided to use Wichman and Hill's 'Psignifit' package (at that time called 'Psychofit') both to fit psychometric functions to my data, and to obtain confidence intervals for their parameters (Hill 1996-2009).

Like Bruce Cumming's program, Psignifit first uses maximum-likelihood estimation in order to determine the best fit between the data and the 'distribution function' used to model it. In maximum likelihood estimation, an algorithm iteratively varies the function parameters in order to minimise the 'maximum likelihood estimator' Θ – a measure of the discrepancy between the sampled psychometric function and values calculated from the distribution function. For my psychometric functions, these are the slope, threshold and lapse-rate.

To determine confidence-intervals for the threshold, Psignifit uses a parametric 'bootstrapping' technique – a form of Monte-Carlo analysis where a computer performs a large number of simulations to generate artificial datasets using the same number of trials per point as the observer's data, and the parameters previously estimated using the maximum likelihood technique. In turn, the parameters of these datasets are estimated using the original maximum likelihood algorithm. The variance of the results is used to indicate the uncertainty in the original estimate that is due to the sampling method and the relatively small number of repeats per point inherent in psychophysics experiments. For relatively small sample numbers, the bootstrapping technique is generally thought to be superior to asymptotic methods such as that used by Watson and Pelli and Bruce Cummings (Hill 2001) or probit methods (Foster and Bischof 1991).

The bootstrapping technique makes two main assumptions: that the observer's actual psychometric function corresponds closely to the function used to fit the data and to generate the artificial datasets, and on the 'bridging' assumption – that the estimated parameters and the actual parameters of the observer's psychometric function are sufficiently close that the variance is similar for both.

Wichmann and Hill performed tests to see whether the validity of the variance estimates provided by their program was seriously compromised by using a function to model the data which was different from the function used to produce it. They found that this could be cause for concern in certain situations, but that the effect was minimal for estimates of slope and threshold when they were being estimated at the 50% probability level (Wichmann and Hill 2001).

To test the bridging assumption, Wichmann and Hill perform additional Monte-Carlo analyses, with threshold and slope values spaced around the original estimate, but using the sample points at which the experiment was conducted. This provides estimates of the reliability of the variance estimates, if there is a significant difference between the parameters of the observer's psychometric function and the maximum likelihood estimates thereof (Wichmann and Hill 2001).

Further discussions of data analysis are provided in the next chapter on statistical inference, and in the experimental chapters where required.

Bias

Conscious and unconscious bias

As an observer in my own experiments, I must remain open to the possibility that I unconsciously biased my responses so that they are more favourable towards the result that I expect, or would like to be true. For this reason, the results of naive observers are more valuable and I therefore discriminate between naive and non-naive observers when reporting the experiments. Every experiment reported here includes results from at least one naive observer.

Bias due to the selection of constants

Studies have shown that the choice of sampling strategy can influence estimates of psychometric thresholds, particularly if the number of trials-per-point is relatively small. Wichmann and Hill investigated a number of sampling strategies, showing that the effects whilst small, certainly exist (Wichmann and Hill 2001).

As previously discussed, in my early experiments reported in this thesis, I chose the constants by hand according to simple heuristics. This approach is clearly suboptimal, and in later experiments I wrote a computer program to place the sampling points automatically, based on the initial threshold estimate.

If sampling points are poorly chosen – whether automatically or by hand – this can cause biased estimates of the threshold. For example, when presented with two alternative choices, observers tend to have a natural with to select each alternative approximately equally often. If one alternative is over-represented in the recent train of stimuli, this may push the observer's measured threshold towards the second alternative (Taylor, Forbes et al. 1983). In a true 2AFC experiment this source of bias should not cause a threshold shift, however in my version it may still be significant.

Colorimetry

Colorimetry is the science of measuring chromaticity – attributes of objects or light sources that we assign to them based on their spectra. Yet colorimetry is not synonymous with the measurement of light spectra – spectrophotometry. As Newton observed, colour is not a property of objects or lights: the term is meaningless except in the context of human perception. Colorimetry is therefore rooted as strongly in the physiology of the human eye as it is in the physics of light.

In its essence, colorimetry attempts to represent light spectra in the same way as the human retina does. The first step in visual processing is that of transduction. The three types of cones present in the retina have peak sensitivities to different portions of the visible spectrum, and the ratio of their output signals is the only information available to the brain about spectral colour distribution. Likewise, the first step in any colorimetric system is to distil the infinity of information available in a spectral measurement down to just three chromatic dimensions, and most colorimetric systems in modern use are based on the CIE 1931 colourspace diagram and primaries.

The CIE 1931 colorimetric system

The Commission Internationale de l'Eclairage (CIE) established in 1931 a set of colour-matching functions which define the appearance of any isolated colour in terms for a standard ('ideal') trichromatic observer. The CIE (1931) color-matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are the three primaries, specified in the wavelength range 380 to 780 nm, of which weighted linear combinations are able to match any light in the visible spectrum. These color-matching functions were obtained from a set of visual matching data from a large number of observers for matching fields of two degrees under photopic conditions. The CIE XYZ trisimulus values of a colour specified by its spectral radiant power distribution $P(\lambda)$ are given by the following equations:

$$X = \int \bar{x}(\lambda)P(\lambda)$$

$$Y = \int \bar{y}(\lambda)P(\lambda)$$

$$Z = \int \bar{z}(\lambda)P(\lambda)$$

Y specifies the photometric luminance of the stimulus, therefore

$$\bar{y}(\lambda) = V(\lambda)$$

specifies the luminous efficiency function. The CIE *xy* chromaticity coordinates are then defined as:

$$x = X/(X+Y+Z)$$

$$y = Y/(X+Y+Z)$$

$$z = (1-x-y)$$ (Wyszecki and Styles 1982).

Common criticisms of the CIE 1931 colourspace amongst life-scientists are that it is neither physiologically based nor intuitive. This is true, but when it was standardised it was not possible to measure the cone sensitivities directly, and it was developed in part with ease of calculation as a goal.

The 1931 colour-matching functions have been subject to revision in recent years, as it is recognised that they are inaccurate for wavelengths below 450nM (Hunt 2001). There are also alternative versions available for a ’10 degree observer’. This is because colour perception for small and large surfaces differs due to a yellow pigmented filter over the fovea, called the macula lutea (Hunt 2001). However, the CS-100 chromameter that I used to calibrate the computer monitor uses the 1931 2^{o} colour matching functions (Minolta), and as far as I’m aware cannot be adapted to different colour matching functions such as the Judd-Voss 1978 2^{o}colour matching functions, updated to be more accurate for short wavelengths (Vos 1978), or the CIE 10^{o} observer,Wyszecki and Styles, Table I, section 3.3.2 (Wyszecki and Styles 1982), though this is probably desirable. Although later measurements taken with the PR-650 Spectrascan could have been used different colour matching functions, the default 1931 functions were used to maintain compatibility.

The practical significance is that there are liable to be calibration inaccuracies in my stimuli, particularly those investigating S-cone isolating stimuli. Tests of S-cone induction are unlikely to be truly isoluminant, and the axis of induction may deviate slightly from the true cardinal axis. Whilst this is regrettable, the 1931 stimuli have proved sufficient for most purposes, and in the context of inter-individual variations due to yellowing of the lens with age and differences in the thickness of the macula (Nayatani, Takahama et al. 1988) they are likely to be accurate enough. Had a greater level of accuracy been critical to my experiments I would have had to create individualised monitor calibrations to take these factors into account.

The MacLeod-Boynton colour-space and cone contrast

Two main types of physiologically derived colour-spaces exist: cone-contrast colour-spaces, and opponent modulation spaces. In my experiments, I use a modification of the MacLeod-Boynton colour-space (MacLeod and Boynton 1979). The MacLeod-Boynton (MCB) colour-space is based around an equiluminant plane in which colours are specified in terms of their cone excitations. The space assumes that only the L and M cones contribute to luminance, and each MacLeod-Boynton plane corresponds to a particular luminance level, i.e. a particular value of $L+M$ (where L represents the L cone excitation of the stimulus, and M the M cone excitation). A colour is specified in this plane by its horizontal and vertical coordinates, l and s, respectively, where l = L/(L+M), m = M/(L+M), and s = S/(L+M). (Note that, as in the original specification of the MacLeod-Boynton space, I use the Smith-Pokorny (1975) cone spectral sensitivities (Smith and Pokorny 1975)).

The total cone contrast of a test stimulus with cone excitations L_t, M_t, S_t, with respect to a background with cone excitations L_{ref}, M_{ref}, S_{ref} is given as (Chaparro, Stromeyer et al. 1993)

$$tcc = \sqrt{\left(\frac{\Delta L}{L_{ref}}\right)^2 + \left(\frac{\Delta M}{M_{ref}}\right)^2 + \left(\frac{\Delta S}{S_{ref}}\right)^2}$$

As is shown above, no correction is made for the fact that the maximal cone contrasts that may be achieved differ greatly between the three axes. This is a drawback common to all similar cone-opponent spaces, and no satisfactory way of resolving it has yet been achieved. For experiments that are confined to a single cardinal chromatic axis this observation is of no consequence, but for intermediate axes it is necessary to compensate for the fact that equal contrasts in the different axes, are not perceptually equivalent. There is no means of compensating for this which is likely to be satisfactory under all circumstances.

In my experiments, I introduce a slightly modified calculation of cone contrast, *mcc*, which incorporates a scaling factor which accounts for differences in luminance between the test patch and the reference background. In most instances, these differences are small and do not appreciably affect the cone-contrast values.

$$mcc = \sqrt{\left(\left(\Delta l/l_{ref}\right) + \left(\Delta m/m_{ref}\right) + \left(\Delta s/s_{ref}\right)\right)}$$

where *l*,*m*, and *s* are the MCB coordinates defined above, and $\Delta l = l_{test} - l_{ref}$. Note that the denominators for l_{test} and l_{ref} may be different if the luminances of the test stimulus and reference background are different.

Testing for anomalous forms of colour vision

There is a wide variation in the degree of colour vision of individuals living in the UK. A complete absence of colour vision is rare, but numerous different forms of anomalous colour vision exist, and many of these are relatively common. Although colour vision seems to be vitally important to primates living in the wild (Onishi, Koike et al. 1999; Jacobs and Williams 2001), the faculty does not seem to be critical to life in the modern world and many colour-deficient individuals remain unaware of their condition. Indeed several artists such as Vincent Van Gogh, Piet Mondrian and Constable were reputed to suffer from anomalous colour vision (Trevor-Roper 1988) – and they seem to have found success despite the visual nature of their work.

Neitz and Neitz reviewed the genetics and mechanisms of anomalous colour vision in humans (Neitz and Neitz 2000). By far the most common inherited deficiency of colour vision is an impairment in so-called 'red-green' discrimination. This occurs because the chromosomal arrangement of the genes that code for the L and M cone pigments is easily perturbed during reproduction, most commonly causing affected individuals to replace one or more of the normal cone types with a hybrid whose properties are intermediate to those of the normal L and M cones. In extreme cases, they may even lack either L cones or M cones, leading to conditions termed 'Protanopia' or 'Deutanopia' respectively. Individuals with these conditions have more extreme deficits.

The L and M genes are situated on the X-chromosome, of which women carry two different copies, and men carry only one. If the genes on a single X-chromosome are damaged, women's colour vision is generally unaffected because their additional copy substitutes; however a man carrying this same affected X-chromosome might exhibit a severe deficit in his colour vision. Such deficits are common: around 8% of Caucasian males have a partial or complete deficit in their L-M colour vision (Fletcher and Voke 1985).

Inherited deficiencies of the S cone are relatively rare as the gene coding for the S cone pigment lies on chromosome 7 (Nathans, Piantanida et al. 1986) and is not subject to the hybridisation issues that plague the genes for the L and M pigments. But the most common acquired deficit of colour vision results from damage to the S cone pathway, as it is physiologically more sensitive to damage than the L and M cone pathway responsible for 'red-green' and achromatic vision. Exposure to toxins such as mercury or organic solvents (Gobba and Cavalleri 2003) therefore tend to cause tritanism over impairments of the L-M pathways. Parkinson's disease is thought to cause tritanism due to the degeneration of dopaminergic neurons in the retina (Muller, Kuhn et al. 1999). Other causes of acquired colour-vision deficiency include inflammation of the optic nerve, which in its early stages causes a generalised decline in chromatic sensitivity that affects both the L-M and S-pathways equally (Mullen and Plant 1986).

An 'anomaly' in colour-vision need not be synonymous with a functional 'deficit'. Due to the phenomenon of Lyonisation, where X-chromosomes are inactivated at random in each cell (Lyon 1961), women may have additional active copies of the genes for 'L' and 'M' pigments. If one or more of these genes is anomalous this may allow the carrier to discriminate colours that are metameric to colour-normal observers. A similar phenomenon is well established in many new-world primates that typically only have dichromatic vision and are only able to distinguish between long and short wavelengths. In some groups of these animals multiple versions of the gene for the L/M-pigment may be found, each with a slightly different spectral sensitivity (Mollon, Bowmaker et al. 1984). Females may carry two different copies of this gene, expressing the different versions in two functionally distinct groups of cone-cells. They appear to be able to compare the outputs of these two cone-types to attain a degree of trichromatic vision , so it seems plausible that human females should be able to attain tetrachromacy by the same means. Indeed, Mollon and Jordan have identified individuals who appear to be functional tetrachromats (Jordan and Mollon 1993).

Even dichromacy may be advantageous in certain situations. Judd suggested that 'colourblind' soldiers should have an advantage in jungle warfare over colour-normal soldiers (Judd 1943), and experiments show that they may be able to detect camouflaged objects more easily (Morgan, Adam et al. 1992). There are also several plates on the commonly used Ishihara test that are designed such that colour-anomalous individuals can interpret them more easily than colour normal individuals (Ishihara 1925). However, in any study of normal colour vision it is obviously important to try and exclude subjects with enhanced or diminished capabilities. Tests for colour vision are discussed in the methods chapter.

Finally, we must ask whether there is such a thing as 'normal colour vision'. Although as previously mentioned, colour vision in primates appears to be highly conserved, variants in the L, M and S-cone pigment genes are known to be present in humans who still have the ability to perform well on basic colour-vision tests (Neitz and Neitz 2000). Of one sample of 62 persons with normal colour vision, 16 variants of the L-photopigment were found (Carroll, Bialozynski et al. 2000). Further, age-related changes in the macula lutea (Hammond, Wooten et al. 1997) and lens (Pokorny, Smith et al. 1987) cause our spectral sensitivity to differ between individuals. The L:M cone ratio is also highly variable. Functional tests show that it can range from around 0.4:1 to 13:1 (Carroll, Bialozynski et al. 2000), and direct visualisation of the photoreceptor matrix has found values between 1.2:1 to 5:1 (Bowmaker, Parry et al. 2003). Although this appears to have surprisingly little effect on the perception of unique yellow (Pokorny and Smith 1985; Yamauchi, Williams et al. 2002), it does influence the luminous sensitivity function V(λ). The practical importance of these factors is that unless individualised, any experimental stimuli can only be accurately calibrated for a hypothetical 'average' observer. Individuals will invariably differ from this ideal, and stimuli that are isoluminant or that intend to isolate functioning in an L-M axis from that in the S-axis, may prove more or less successful for different observers depending on their individual form of colour vision.

The Farnsworth-Munsell 100-hue test

The Farnsworth-Munsell 100-hue test is a well established colour vision test, in which observers are asked to arrange painted chips, so that their colours form a smooth progression (Farnsworth 1957). After the observer has done this, the experimenter checks the order of the chips by upturning them and reading the numbers underneath. Out-of-order chips result in a higher score.

The mean score obtained in the 100-hue test only provides a measure of how good an individual's colour discrimination is; taken alone it does not tell us whether they have an anomalous form of colour vision. To determine this, it is necessary to examine the pattern of mistakes that they make in a somewhat subjective manner. However, as we are using the test as a form of screening rather than as a diagnostic test, there is no need for us to perform more rigorous investigations such as Nagel-anomaloscopy (Williams 1915).

Flicker photometry and minimum-motion photometry

Aside from differences in the sensitivity to chromatic variation, the luminous efficiency function may differ considerably between observers. Flicker photometry provides one way of measuring these differences, and although a CRT monitor is only able to measure part of the luminous efficiency function, it measures precisely that part which is necessary in order to generate isoluminant stimuli using the CRT monitor in question.

In flicker photometry, two phosphors (e.g. red and green) are rapidly alternated, and the observer's task is to adjust their relative brightnesses until the perceptual salience of the flickering reaches a minimum. This provides a measure of the isoluminant point for these two phosphors for that particular monitor, for that particular observer (Kelly 1983; Ladunga 2001). Minimum-motion photometry works on a very similar principle: chromatic sine-waves are produced using the red and green phosphors of the monitor, and set into opposing directions of motion. The combined percept has a composite motion that depends on the relative brightnesses of the two phosphors. By adjusting these, the observer can attempt to eliminate this motion percept and thereby find the isoluminant setting (Anstis and Cavanagh 1983).

A necessary caution is that the results obtained by both techniques are dependent on the temporal and spatial characteristics of the stimuli, so an isoluminant match made at 1 Hz may no longer be isoluminant at 8 Hz. Cavanagh briefly details some of the difficulties inherent in the techniques (Cavanagh 1991), and suggests that rather than risk picking an incorrect equiluminance setting, experiments should be repeated several times at different settings so that the results for the most equiluminant settings should be identified.

Although I used flicker-photometry on some pilot experiments not reported in this thesis, I did not use it on a regular basis for several reasons. Firstly, the stimuli were unpleasant to look at, causing me to feel slightly nauseated. I was therefore concerned about the risk of causing epilepsy in photosensitive individuals. This was probably overcautious – the yearly incidence of new diagnoses of photosensitive epilepsy in 7-19 year-olds with no previous history of epilepsy is in the region of 1.5/100,000 (Quirk, Fish et al. 1995). Nonetheless, stimuli with a chromatic component have provoked isolated attacks in individuals who were not otherwise known to be susceptible (Tobimatsu, Zhang et al. 1999), although the stimuli implicated generally appear to have a strong S-stimulating component which is not the case for flicker photometry stimuli. Secondly, producing individualised stimuli would also have added considerable complexity to my experiments. But finally, my experiments for which it was most important to find an isoluminant setting showed that there was little interaction between isoluminant and chromatically defined stimuli, meaning that I had already achieved an adequate degree of isoluminance.

Apparatus

One of the attractions of psychophysics is that elegant experiments can often be performed with relatively modest equipment. In recent years, psychophysicists have increasingly come to rely on computers – to generate psychophysical stimuli, to collect responses and then to analyse the data. My experiments were entirely computer based, but I additionally used a colorimeter and spectrophotometer to calibrate the monitor. Paper-lined viewing boxes were used to control adaptation, and a Wheatstone stereoscope was used to produce stimuli in stereoscopic space. All experiments were performed in a light-tight psychophysics tunnel.

The psychophysics tunnel

A laboratory designed for experiments into visual psychophysics is known as a 'psychophysics tunnel.' As my experiments were computer-based, our psychophysics tunnel simply consisted of a light-tight room with a computer monitor, viewer and input device such as a gamepad. The observer sat in a chair, resting his head against a chinrest in order to fix his head position relative to the computer monitor and holding the input-device. For pictures of our setup, see **Error! Reference source not found.**.

Light-tightness

In preparing a psychophysics tunnel, there are two main aims. Firstly, it is important to ensure that unwanted light-sources are obscured; secondly, surfaces should be painted a dark colour so that they reflect a minimum of light from legitimate sources of light such as the computer display.

A prerequisite for most psychophysics experiments is the ability to control each relevant aspect of stimulus presentation. Most importantly for my experiments, I need to have complete control over what the observer is able to see, avoiding the possibility that their responses might be influenced by extraneous objects or light sources, such as white walls, coloured LED indicators on computer equipment, light penetrating from adjoining rooms or under doors, or similar unwanted stimuli. I also wished to ensure that my observers could attain a stable state of adaptation – which might be broken if they were to look back and forth to irrelevant sources of light.

Safety

It can be dangerous to walk about in an unfamiliar, darkened room. I have always taken care that my observers have had a desktop lamp placed within easy reach, so that they could exit the tunnel safely in case of a fire alarm or if they have any other reason to do so before an observing session was complete.

a)

b)

c)

Figure 1: Scenes from the tunnel

a) *A communications port, to allow a computer in an adjacent room to control the experimental monitor in the tunnel. The walls of the room are painted a dark colour to minimise light reflections.*

b) *Monitor, keyboard, gamepad and light. Note the black tape placed over the keyboard indicator lights, the lamp for use in emergencies, and the dark-coloured surfaces.*

c) *A draught-excluder at the base of the door prevents light from entering underneath. Note the rubber beading around the inside and outside edges of the door-frame. This prevents stray light from entering the room by passing round the edges of the door, or by passing through cracks in the door-frame.*

Computers and Psychophysics Experiments

Computers are widely used in visual psychophysics as they allow the automation of data collection; they can potentially generate accurately controlled, complex stimuli; and they erect an impersonal barrier between experimenter and observer. I believe this last factor reduces the likelihood of the experimenter unconsciously influencing the observer's responses. I used off-the-shelf PCs and a Silicon Graphics workstation to run my experiments, and switched between various operating systems including Windows 98, Windows XP, Linux and Irix. My experiments were not demanding of the processor or hard disk, but they sometimes required large amounts of memory or an accelerated graphics card.

The main components of a computer system for use in psychophysics experiments are a computer, graphics card, display device and an input device.

Operating system & other software

In my opinion, the choice of operating system is largely one of personal preference, which is not to say that modern operating systems are eminently suited for running psychophysics experiments. One major problem is that newer operating-systems support multi-tasking: they allow the processor to switch rapidly from one program to the next to create the illusion that several programs are running concurrently.

For psychophysicists, this can create some very real problems (Kleiner 2003). Multi-tasking creates 'glitches' that may interfere with experiments that depend on accurate time intervals. Perhaps a particular stimulus was only meant to be displayed for 100 milliseconds before being masked by a random dot display. A multi-tasking computer could conceivably write the stimulus to the screen then spend 200 milliseconds checking for email before returning to the psychophysics program to update the display. Such are the potential pitfalls associated with multi-tasking that some experimenters choose to use older operating systems such as the Microsoft DOS, which only run a single program at a time. Or they may use dedicated display drivers such as Cambridge Research Systems Visual Stimulus Generator cards (ViSaGe) – essentially a dedicated display-driving computer combined with a video card that does not run a multi-tasking operating system.

A third method, particularly well suited to trial-based experiments, is to use a multi-tasking operating system but to monitor the experiment timings very closely. If any trials are adversely affected by glitches, this is detected and the results of those trials discarded. When appropriate these trials may be repeated later in the experiment. In practice and provided the stimuli are relatively simple, little data is likely to be lost and the experimental programmer benefits from the advanced features of modern operating systems and graphics cards.

This is the method used by the popular psychophysics stimulus control program *Presentation (Neurobehavioural Systems, 2001)* that I used for some of my experiments, and it is also the strategy I chose when I programmed my own trial-based experiments from scratch. It is less well suited to experiments that use lengthy animations such as my motion-contrast experiments. There were very few, provided that superfluous programs (web browsers, emailers etc.) were stopped before running the experiments. Appendix G describes a phototransistor device I designed and built, and discusses how I used it to verify the timings of my experiments.

Graphics cards

Graphics cards perform two fundamental functions. They store a digital copy of the screen contents in video memory, termed the '*active framebuffer*', and they convert it to a video signal that is output to a computer monitor or projector. A good graphics card should have enough memory to store at least two separate framebuffers so that one (the 'back buffer') can be updated whilst the active framebuffer is being displayed. This is important as severe artefacts may occur if the framebuffer is updated whilst the video card is drawing its contents to the screen. 3D graphics cards with more advanced features such as texture-mapping and anti-aliasing are also available, and these features may be necessary if complex stimuli are to be generated in real-time. Cards that support higher resolutions and refresh rates require higher quality components with a better high-frequency response. Even at low resolutions, they should produce a higher-quality video output than less expensive graphics cards.

Response collection

a) b) c)

Figure 2: Examples of input devices that I have used in my experiments

a) A custom-built 'button-box'

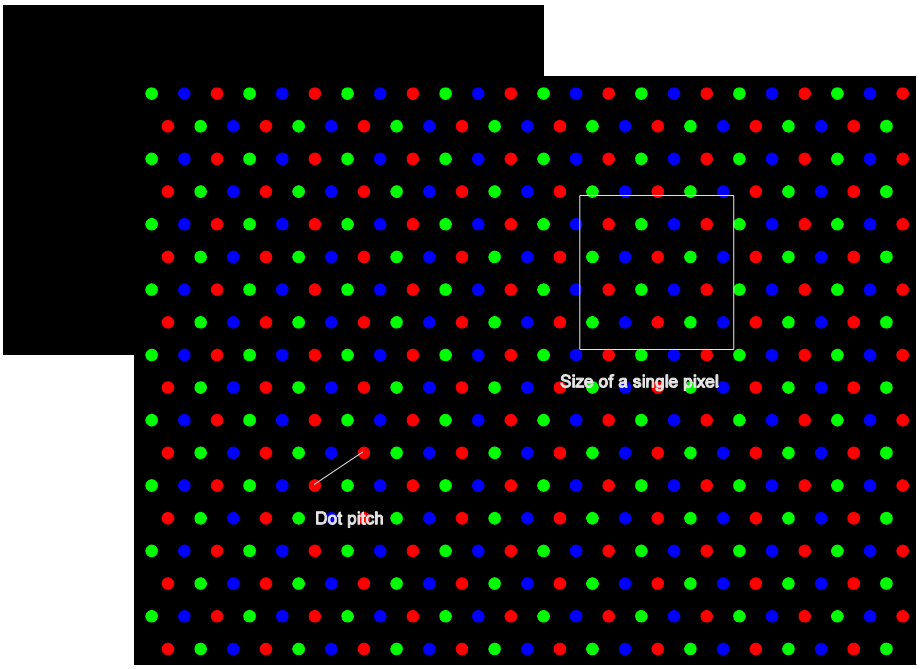
b) Mouse

c) Gamepad

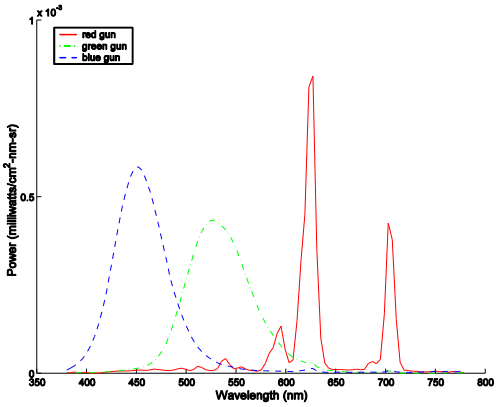
My psychophysics experiments are performed in the dark, which can make it difficult to find the keys on computer keyboards. In a few experiments, I glued pieces of paper to the keyboard to act as tactile cues. But for the most part I used either mouse-buttons, a gamepad or a custom-built button-box as input devices, shown in figure 2. The gamepad is probably the most comfortable to use: an important consideration during long experiments. Details of the input devices used are recorded in the experimental chapters in the descriptions of each experiment.

Before starting each session, I ran through the experimental procedure and button assignments with the observers. If desired, they could run a short practice experiment (whose results were discarded) to ensure that they understood how to carry out the task.

CRT monitors



a)

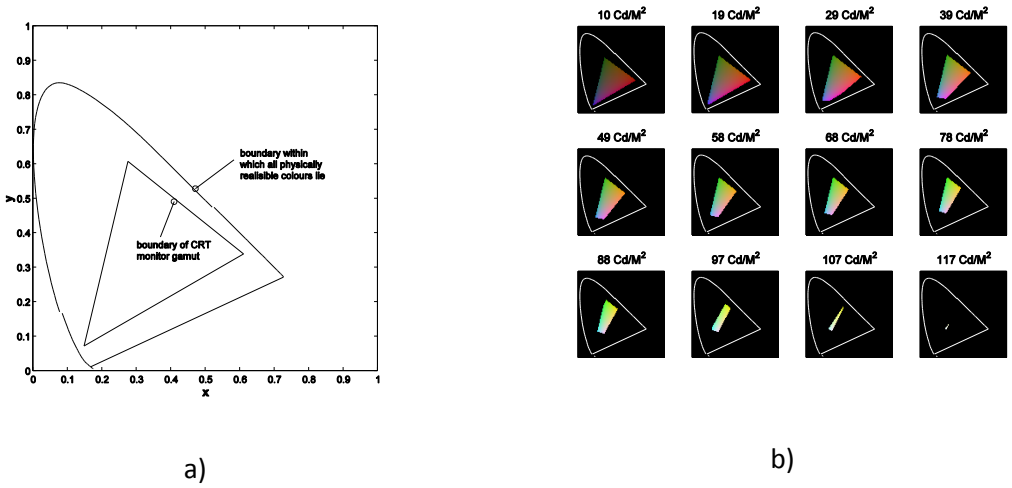


b)

Figure 3: The arrangement and spectra of the three phosphors on a CRT monitor

a) The array of phosphor patches of each colour are arranged for many CRT monitors. The 'dot-pitch' on each monitor determines the smallest pixel that may be displayed, as each pixel must include at least one patch of phosphor of each colour. However, pixels may also be larger than is decreed by the dot-pitch- the white square illustrates the area that might be covered by a typical pixel.

b) Typical phosphor spectra for each of the three primaries.



a) b)

Figure 4: The gamut of a CRT monitor

a) A chromaticity diagram (CIE 1931) showing the boundary that encloses all physically realisible colours, with the gamut of a standard CRT monitor superimposed. Many colours found in nature cannot be reproduced on any standard computer display.

b) These 12 diagrams show how the monitor gamut shrinks with increasing brightness.

Most (though not all) computer displays consist of an array of small points, known as 'pixels' – whose colours may be varied independently of one another. If the pixels are too fine for the human eye to resolve individually, such a display may be used to present a smooth 2-dimensional image.

The screen on a colour cathode-ray-tube based display, or 'CRT' consists of a thick pane of glass, coated with a fine array of three different phosphor mixtures as shown in figure 3a. Each mix produces a different coloured light when it is bombarded by an electron beam, but the phosphor array is typically very fine, so the eye is unable to resolve the individual coloured points that comprise a single pixel. As discussed in the section on colorimetry, it is possible to produce different colours by mixing three well chosen ‘primaries’, and at normal viewing distances such optical blurring accomplishes this effectively.

Whilst any colour may be matched by adding or subtracting three well-chosen primaries, a CRT monitor can only add primaries together, so the more saturated colours will lie outside of its gamut. Those colours that can be displayed all lie within the boundaries of the triangle formed by plotting the three primaries – the phosphor colours – on the CIE 1931 chromaticity diagram (figure 4a). Because the maximum brightness of the red and blue phosphors is less than that of the green phosphors, CRT screens display can only display most saturated colours relatively dimly; the gamut contracts as the brightness is increased (figure 4b) and the full gamut may only be available at settings less than 10% as bright as the brightest white that can be displayed.

Artifacts of computer monitors

Phosphor persistence and temporal frequency response

Different phosphor mixes continue to glow for different durations, after they have been excited by an electron beam. Their brightness decays exponentially, so their persistence is given in terms of a 'half-life' that may range from a few microseconds to many milliseconds. In general, red phosphors have a long persistence of around a millisecond. Green and blue phosphors vanish much faster, with persistences in the region of 30 to 50 microseconds.

The persistence of the red phosphor can cause problems in certain applications – particularly when stereo goggles are used with shutters that allow the left and right eyes to see alternate frames. Phosphor persistence may cause the bottom part of the last frame to be visible, at the same time as the top half of the next frame is presented to the next eye. However, even with a fast refresh rate of 100 Hz, each frame lasts for 8-20 half-lives. Even in a worst-case scenario, the intensity of the red-phosphor will fall to less than 1/256 of its original level within the duration of a single frame. For this reason, I do not believe that this artefact is relevant to my experiments. It is safe to say that I can control the display of any single part of the screen with a temporal resolution of a single frame-duration.

Temporal dependence

I suspect this artefact is not directly related to phosphor persistence, but is probably due to thermal effects in the electron-guns or control-electronics. Its time-course is much longer than phosphor persistence, and it occurs to different degrees in different makes and model of monitor – its magnitude ranges from highly significant to unmeasurable.

This artefact may be elicited by displaying a bright red screen, with a small yellow patch (a combination of red and green) in the center. Over the course of the next half hour, the brightness of the red phosphor typically diminishes but the brightness of the green phosphor remains constant. The result is that the display gradually becomes slightly dimmer, and the yellow patch becomes gradually greener. I avoided using monitors that strongly exhibited this effect, and I believe it could be a significant problem in experimental designs that require long-term stability, such as experiments investigating adaptation to coloured backgrounds. However, individual trials in my experiments never lasted more than a few seconds or involved long-term colour memory - so I do not believe such a problem is likely to have affected them significantly. However, I have not investigated the possibility that the gamma function – the relationship between input and output brightness – changes due to this artefact.

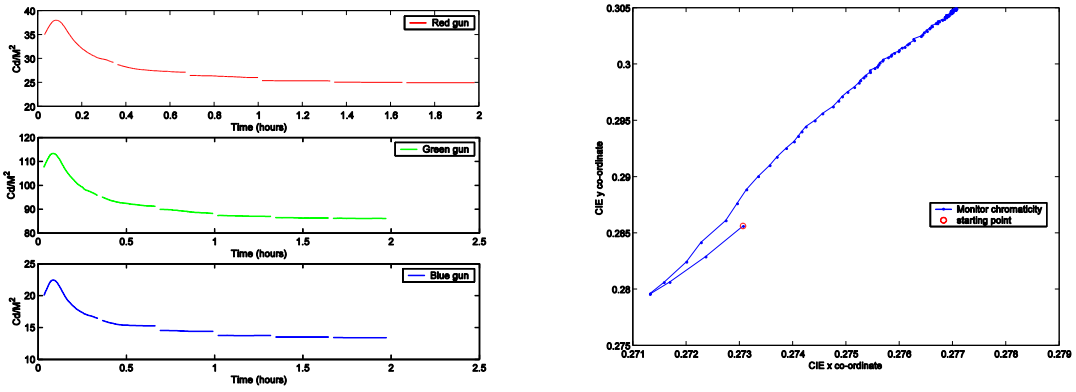
I do not believe this form of temporal dependence is related to heating of the phosphors, because they are mounted adjacent to each other on the glass screen of the monitor. If one phosphor became hotter than the others, this heat would very quickly be transmitted to adjacent phosphors and all would be affected.

Temporal instability

It is only possible to be absolutely sure that a monitor's calibration is correct, at the moment it is calibrated, and at any subsequent time at which it is tested. For all intermediate points, the experimenter can only work on the assumption that the monitor's calibration has not drifted.

Temporal instability may conceivably take place on several scales, that may be related to the main causes of variation in monitor calibration: namely ageing, which generally takes a long time – and thermal factors, that may vary on much shorter timescales, even down to the duration of a single experimental session.

Figure 5 shows how the brightness and neutral-colour of the 21" HP monitor vary during the first half- hour after it is turned on.

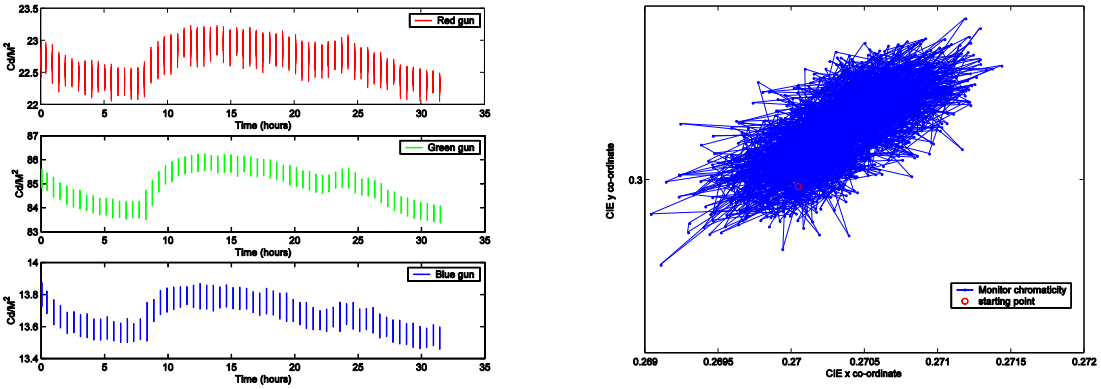


a) b)

Figure 5: Instability during the 'warmup' period

1. *Immediately after the monitor is switched on, the brightness rises by around 10% before gradually dropping until it is 10% dimmer than its starting level. After an hour, the brightness variation approaches its asymptote.*
2. *During the first hour, the neutral-chromaticity of the monitor varies measurably but for most purposes, probably not significantly. The blue line shows the chromaticity of a neutral white. The solid dots punctuate intervals of 1 minute's duration and the red circle indicates the starting chromaticity. After an hour or so, the monitor chromaticity has largely settled.*

Figure 6a shows the drift in monitor calibration over the course of a day and a half after it has already been warmed up. It is surprisingly small – the brightness of the red primary changes at most by 1%. Note that the brightnesses of the three primaries change roughly in unison, meaning that chromaticity variation is minimal. Figure 6b shows the variation in chromaticity of a neutral white over this period this time – it is insignificant.



a) b)

Figure 6: 'drift' in monitor output characteristics over 34 hours for the 21" HP CRT monitor used in many of my experiments

a) Variation in brightness for the red, green and blue primaries. The standard deviations of these readings are 1%, 0.85% and 0.66% respectively – note that the Y-axis scale is much smaller than in Illustration 7. Part of a 24-hour rhythm can be discerned: this is probably due to the lab-heating switching on and off.

b) Variation in chromaticity: standard deviation in the 'x' axis is 0.00038; standard deviation along the 'y' axis is 0.00045; the deviations in chromaticity are negligible

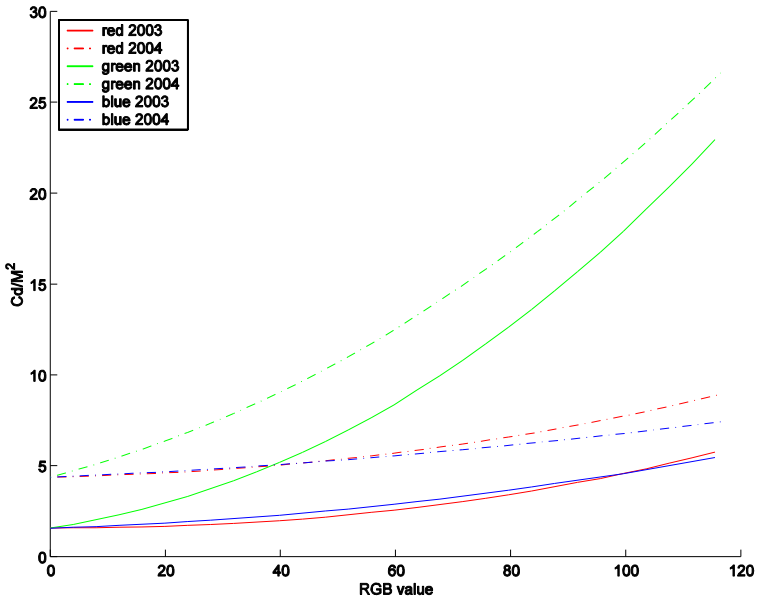


Figure 7: Variation in 'baseline glow' over time

However, over long periods of time, the monitor's characteristics may change considerably. For example, between June 2003 and June 2004 the HP monitor's baseline glow increased from 1.5Cd/M^2 to over 4 Cd/M^2 (figure 7).

In an experimental setting, long-term and 'warmup' drift appear to be the most important problems to be aware of. For this reason I left the monitor running most of the time, and tried to power it up an hour to half an hour before starting any experimental sessions in the morning. However, I did not recalibrate the monitor during runs of experiments, as I felt it would be better to combine results that were obtained with the same slightly erroneous calibration, than to try and combine results that were obtained with two separate monitor calibrations.

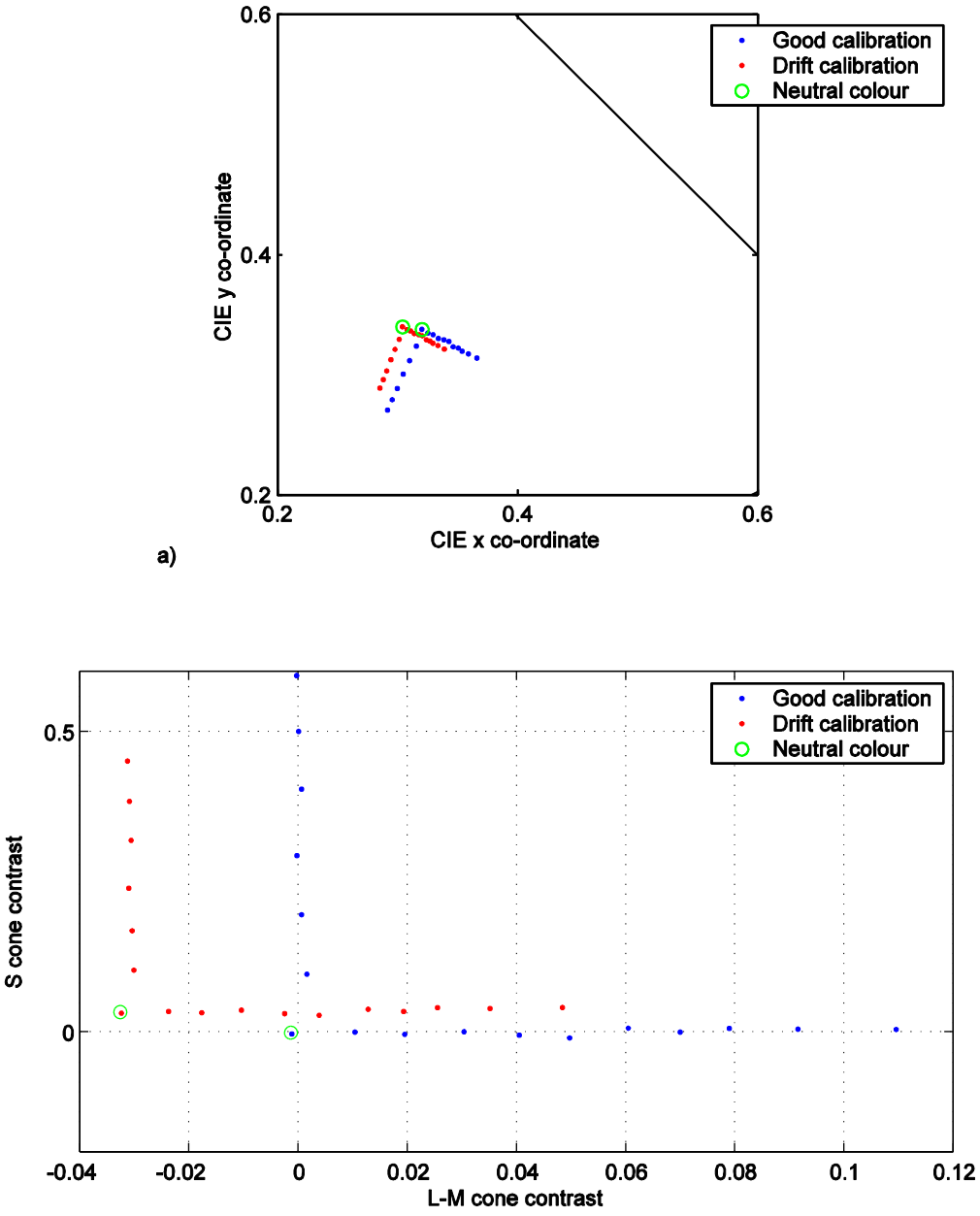


Figure 8: Effect of monitor-drift on a set of chromaticity co-ordinates. The sets consist of a range of constant-stimuli lying both along the L-M axis, and along the S-axis.

a) Shows the co-ordinates plotted on the CIE 1931 chromaticity diagram

b) Shows the co-ordinates plotted in a modified McLeod-Boynton colour-space (whitepoint [30, 0.321, 0.337])

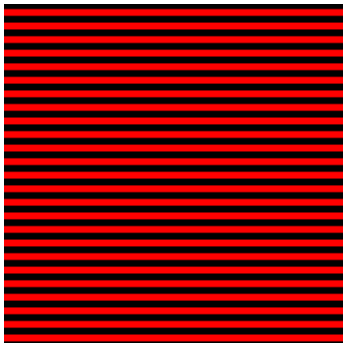
When plotted in the McLeod-Boynton colour-space, it is clear that the co-ordinates have been translated and rescaled. However, the directions of the chromatic axes have not changed. The scaling factors for each of the dimensions are within 10% of each other for the two chromatic dimensions and within 20% for the brightness-dimension relative to the chromatic dimensions. The brightness dimension is not shown on this diagram.

Figure 8 shows how the colours on a monitor may change if it drifts far out of calibration. The significance of these errors depends very much on the experiment that is being run, though of course it would not be 'good-practice' to run experiments on such an ill-calibrated monitor.

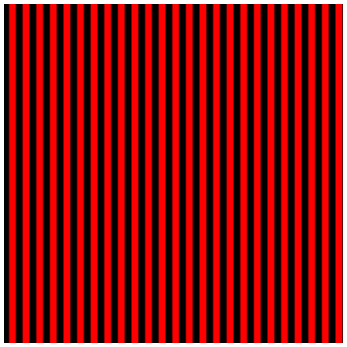
The 'neutral colour' or origin of the pattern is translated quite far – which would be very important in an experiment to measure long-term colour-memory. However, in my experiments, the colour of the neutral point is arbitrary, and observers are never asked to memorise any colours for longer than a few seconds. The unequal rescaling along each is of more concern, however it is gratifying that the figure is not rotated relative to the chromatic axes as my stimuli are tightly tuned to them.

The modulation-transfer-function

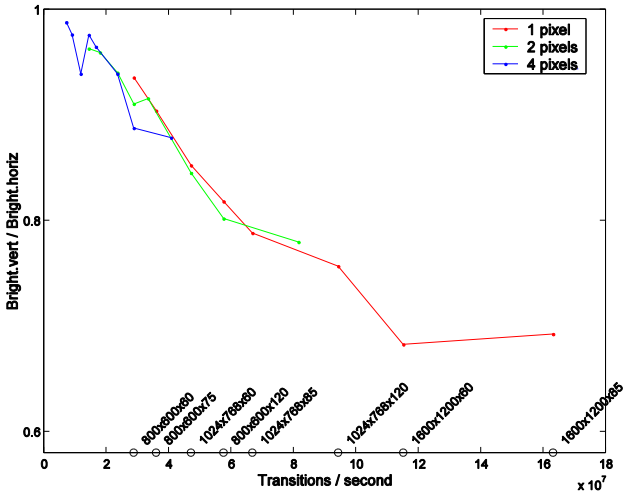
How faithfully is a pattern stored within the framebuffer, converted into an image on the computer screen? Monitors are typically characterised by displaying a large coloured patch on the screen, whose colour can be measured by a spectrophotometer or colorimeter. However, the stimuli used in psychophysics experiments may contain elements at many different spatial scales – some of which are too small to be measured by our instruments. How can we be sure that our stimuli are are accurately drawn, when we cannot measure them directly?



a)



b)



c)

Figure 9: Testing the monitor's modulation-transfer-function

a) Horizontal grids used to test the monitor's low-frequency response

b) Vertical grids used to test the monitor's high-frequency response

c) The brightness of the vertical grids divided by the brightness of the horizontal grids, plotted against the number of discrete pixels drawn per second. This graph shows that the artefact is most pronounced for high screen resolutions with high refresh rates. Lower pixel rates were simulated by drawing lines that were 2 or 4 true pixels in width.

Mollon (Mollon 1999) points out that intensity of the beam from an electron-gun is exponentially related the voltage of the video signal, so if a signal containing a bright and a dark colour is averaged in voltage, the resulting colour will be considerably less bright than the spatial average of its bright and dark component colours would have been. This provides an indirect method of investigating the modulation-transfer-function at fine spatial scales. Figure 9a and b show grids that should appear identical, save for the fact that one is oriented vertically and the other is oriented horizontally. However, they are very different in terms of the raster pattern required to generate them. As the raster pattern traces across the vertical grid, the electron gun must switch on and off at a very high frequency. But as it draws the horizontal grid, it need only change

once for each horizontal line. By using the horizontal grid as a reference and comparing it to the dimmer vertical grid, we can estimate the modulation-transfer function at different gun-modulation frequencies.

Figure 9c shows the relative overall-brightnesses of the vertical and horizontal grids for the HP-21” monitor used in the majority of my experiments. If the framebuffer were to be represented faithfully on the screen, the brightnesses of the horizontal and vertical grids would be identical (a value of '1' on the Y-axis). If the averaging was near-complete, the value on the Y-axis would be around 0.25.

The data here represent the errors that result during 100% modulation – clearly a worst-case scenario. It is difficult to predict the exact effect that this artefact may have on my experimental stimuli as they do not contain such high contrasts, but it is worth noting that errors are much reduced at the lower resolutions and moderate refresh rates that I used in my experiments. These differ from experiment to experiment, and precise details are given in the individual methods sections and experiment-specific appendices, but the highest resolution that I used was 1024 x 768 pixels at 85 Hz.

Spatial inhomogeneity

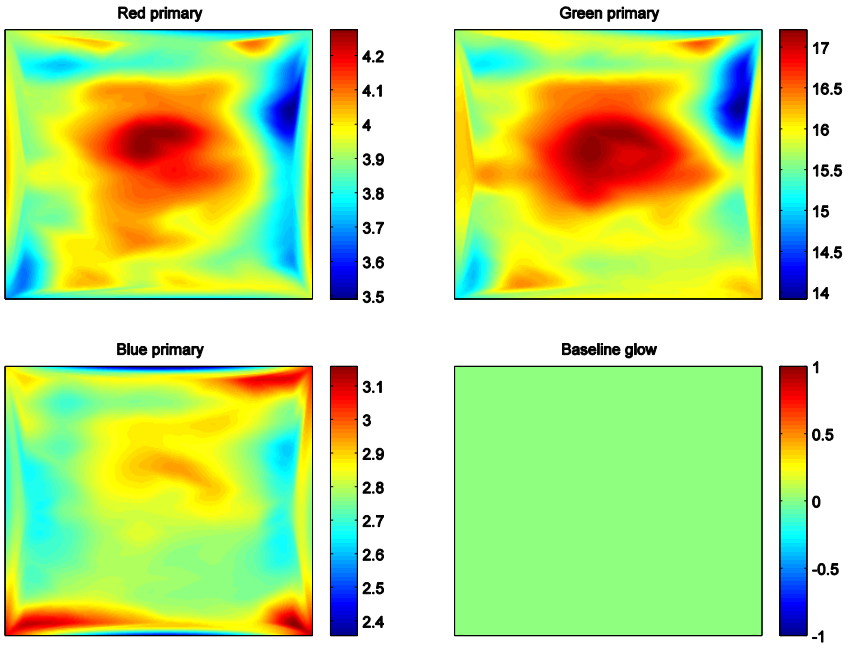


Figure 10: Brightness across the screen for each of the three primaries and for the 'baseline glow' - expressed in Candelas-per meter squared for RGB level [200, 200, 200]. The monitor was an SGI 20" curved screen model.

As the yoke directs the electron-beams to different parts of the screen, it can also cause them to vary in brightness (figure 10). This can lead to a fall-off in brightness across the screen. Thankfully, the red, green and blue guns are typically affected to the same extent and the gamma-function does not change (figure 11), so although some monitors exhibit variations in brightness of up to 20%, large changes in chromaticity do not occur.

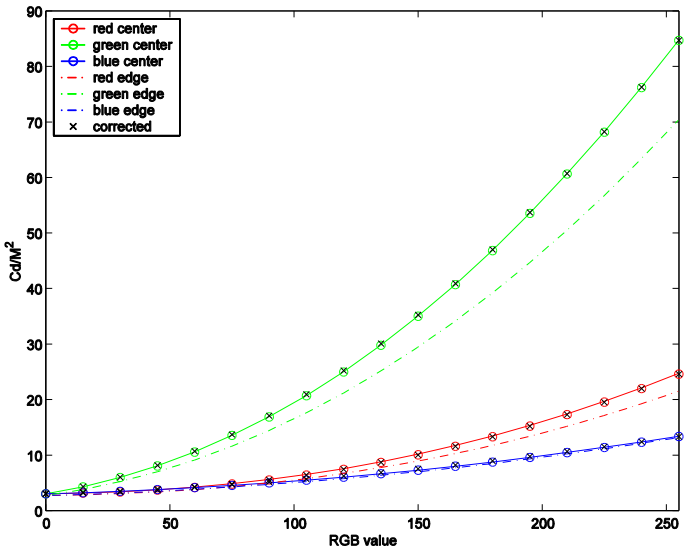


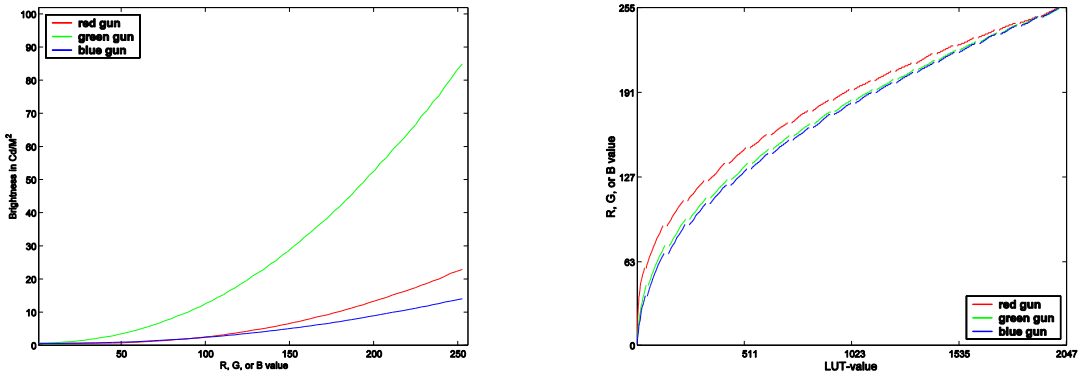
Figure 11: The gamma function at the center of the 21" HP monitor's screen (solid lines with circles) and at the dimmer edge of the display (hash lines). A simple multiplicative correction causes the edge and center curves to overlap, showing that the exponent of the gamma function is identical at both parts of the screen.

Characterising a CRT monitor using a simple additive model

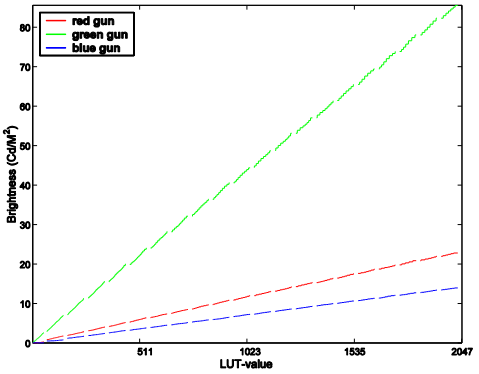
In section 2.2.1 I discussed the fact that three well-chosen primaries may be mixed in order to reproduce the appearance of a colour-signal with any given spectrum – subject to certain limitations. Each pixel on a CRT monitor is composed of three phosphors – one red, one green and one blue. The goal of CRT characterisation is to vary the brightnesses of these three primaries accurately and independently for each pixel, thus building up a 2D image on the screen.

Gamma correction using lookup-tables

Current video cards generally contain three 8-bit analogue-to-digital converters offering a total resolution of 256 settings per electron gun. The setting on each gun is thankfully largely independent, as has previously been discussed. However, the relationship between the digital input and the phosphor brightness is non-linear as is shown in figure 11a, so the input-output relationship of the monitor, termed its 'gamma function', must be fully characterised before the experimenter can control the brightness to order.



a) b)



c)

Figure 12: Creating a linearised lookup table

a) A CRT 'native' gamma function

b) The inverse gamma function

c) A plot of the linearised, expanded lookup table

The relationship between the lut-values and their corresponding RGB values is highly non-linear, as shown in illustration a. However, the relationship between a cell's position in the lookup table and the brightness of the RGB value that it represents is linear (illustration c) and corrects for the monitor's non-linear gamma function.

Whilst monitor gamma functions can be described mathematically, many psychophysicists use linearised lookup tables (LUTs) to choose the digital setting that will result in a specified brightness. A table is prepared such that first entry contains the digital value corresponding to zero brightness; the value corresponding to the maximum possible brightness resides in the last cell and intermediate values are chosen such that they will output a brightness proportional to their position in the table. To find the R, G or B value corresponding to 25% of the maximum brightness for that primary, the experimenter looks up the value a quarter of the way between the first and last entries (see table 1 for an example).

Calibration program

As part of a larger suite of programs ('kccv') which I wrote to carry out conversions between colour space, I created calibration software which interfaces with a chromameter (Minolta CS-100). (See Appendix C). My code uses an expanded lookup table with 2048 entries per gun mapping to 256 possible output levels – this is wasteful in that most output levels are represented several times in the lookup table, but conservative in that almost all are represented at least once. This ensures that little or no output resolution is lost.

LUT entry	'G' value (RGB)	Luminance
0	0	0 CdM^{-2}
…	…	…
511	146	22.5 CdM^{-2}
…	…	…
519	146	22.5 CdM^{-2}
520	147	22.75 CdM^{-2}
…	…	…
2047	255	90 CdM^{-2}

Table 1: The relationship between the expanded lookup table, the RGB values used for the green gun, and the brightness of the monitor specified in Candelas per Meter^2. Individual RGB values may be represented several times in the lookup-table, as is illustrated by RGB level 146, which occupies LUT entries 511-519. Note that this LUT has been corrected to start from a brightness level of '0 Cd/M^2'. In real life, all display devices produce some stray light – even when the whole screen is set to display RGB value [0, 0, 0].

'Baseline glow' correction

A computer monitor is never absolutely black, even at the dimmest setting (RGB 0, 0, 0). This leads to errors in calibration at low output levels unless a CRT characterisation model takes this into account. For example, the brightnesses of RGB levels [1, 0, 0], [0, 1, 0] and [0, 0, 1] may be as much as 1 Cd/M^2. A basic additive model might predict that the brightness of RGB colour [1, 1, 1] would be 3 Cd/M^2 but in practice it would not be significantly different from RGB colour [1, 0, 0]. Jiménez del Barco et al. (1995) (L. Jiménez del Barco 1995) modelled the output of a monitor as an additive combination of a constant baseline glow and a variable contribution from the red, green and blue guns(L. Jiménez del Barco 1995). My calibration code also uses this method to improve accuracy, especially for those colours with lower RGB values.

Halftoning

Unfortunately, human colour vision can be so sensitive, that we can sometimes detect the difference between RGB levels that differ by only a single value. Video cards are available that provide increased chromatic resolution – for example Cambridge-Research systems market a custom-designed stimulus generator that has 14-bit digital-to-analogue converters to improve the colourspace resolution (Anon 2006). But a cheaper and more adaptable method is to use halftoning to increase the chromatic resolution of the stimuli (Robson 1999).

'RGB' setting	***Luminance***
[0, 146, 0]	*22.5 CdM^{-2}*
[0, 147, 0]	*22.75 CdM^{-2}*

Table 2: Luminances of two adjacent entries for the brightness of the green primary for a typical CRT monitor

The technique of halftoning trades spatial or temporal resolution for increased chromatic resolution. Table 2 shows luminance levels for two adjacent RGB settings – the luminance increment between the upper and lower level is in the region of 1% which is relatively large. To produce intermediate steps, the relative proportions of two adjacent RGB levels may be varied over a small area. Provided that the eye is insensitive to this low level of contrast-variation at the spatial scale used for the halftoning pattern and that the spatial features of the stimuli are large in comparison with the scale of the halftoning pattern, the result is equivalent to an uniform coloured patch with the same space-averaged luminance and chromaticity.

In experiment 1, I used halftoning to double the chromatic resolution of my stimuli. To do this, I determined the two RGB triplets whose mean was closest to the chromaticity and luminance that I desired (they could be identical), and coloured even and odd horizontal lines using the first and second RGB values respectively.

In later experiments, the shifts in colour appearance that I was investigating were far greater, so the precision with which I needed to manipulate my stimuli was correspondingly smaller. For simplicity's sake, I did not use any form of halftoning in these experiments.

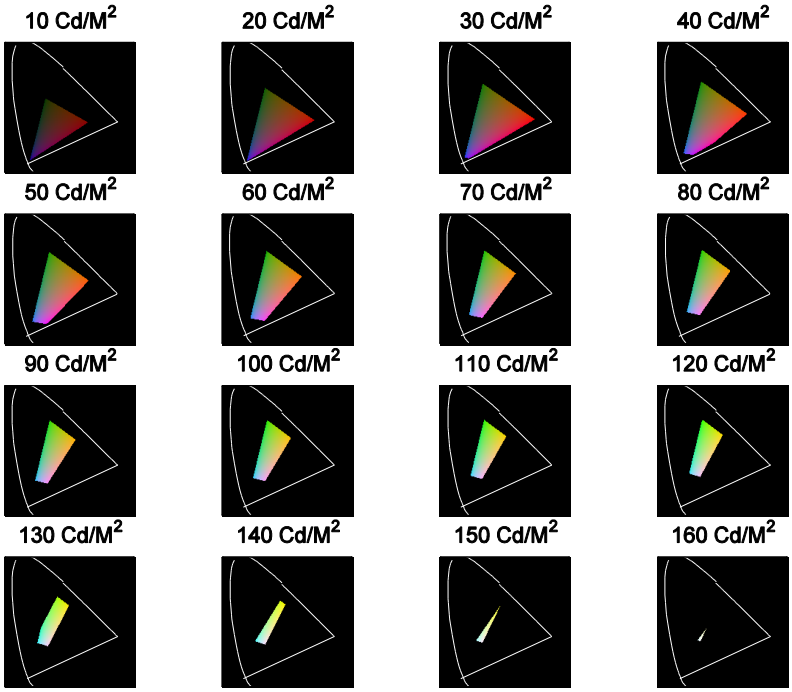


Figure 13: monitor gamut relative to the CIE 1931 chromaticity diagram, at luminances between 10 and 160 Cd/M²

The gamut available for colour-experiments is constrained primarily by the choices of the three primaries; however it also varies as a function of the brightness of the colour being displayed. Figure 13 shows the differing gamuts at luminances between 10 and 160 Cd/M^2 for the 21" monitor.

It is generally desirable for experiments into colour vision to increase the luminance of the stimuli to as high a level as possible, to avoid rod input at mesopic levels and to increase the finesse with which chromaticity may be varied. However, as brightness is increased the gamut decreases and a compromise must be chosen.

As the blue phosphor has the lowest luminance, it sets the maximum brightness at which the full gamut of the monitor may be realized.

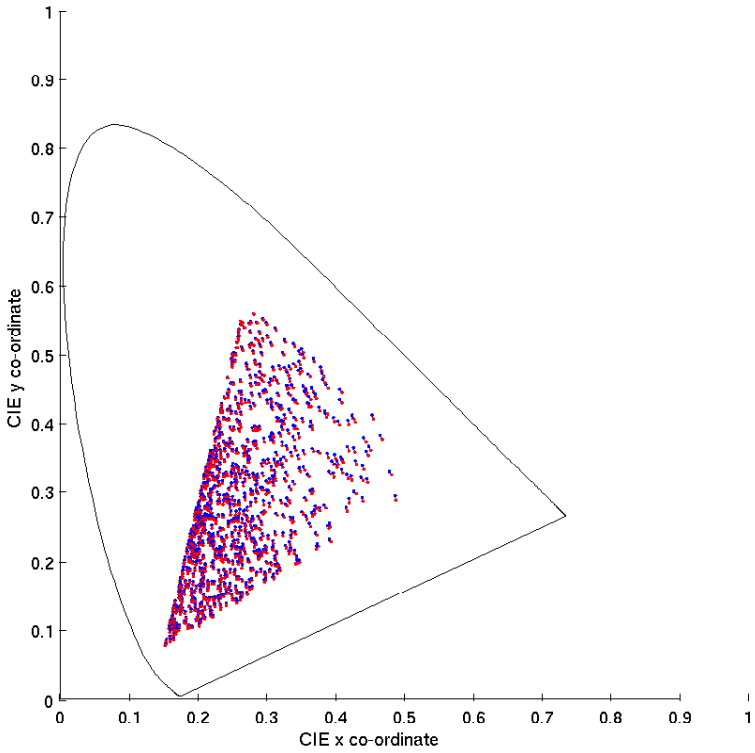


Figure 14: Predicted (blue) and measured (red) chromaticities of randomly generated colours show a small but significant bias in the monitor characterization.

To determine the accuracy of the generated colours, I wrote an additional program as part of the 'kccv' suite that generates random combinations of RGB values, predicts their Yxy co-ordinates (see section on Colorimetry below), then compares these values to those measured by the spectrophotometer or spectroradiometer. Figure 14 shows blue dots (predicted chromaticities) overlaid by red dots (measured chromaticities).

Chapter 4 - Null-Hypothesis Significance Testing

The methods used to fit psychometric functions generally provide a number of parameters that specify the shape of the modeling function, and many of them also estimate confidence intervals for these parameters. However, these descriptive statistics are rarely useful if studied in isolation. In general, it is also necessary to compare them to the parameters of other psychometric functions in order to make any useful statements about what an experiment shows. But if two measured functions do appear to differ, how can we have confidence that this is genuine and not simply due to sampling error?

The methods of inferential statistics, in particular 'null-hypothesis significance testing' (NHST), provide one means of approaching this problem. However, as shown in Appendix D, inferential statistics have only recently started to be commonly used in the analysis of psychophysics data. My own view is that this is at least partly because their application to this field is not at all straightforward. In this chapter, I discuss the aims and interpretation of significance tests, before discussing whether significance-testing is actually necessary or desirable for my data, and I overview some alternatives and adjuncts that have been proposed. More detailed discussions of methods are provided in the appendices.

The aims and interpretation of significance tests

The present-day concept of statistical significance testing was developed in the first half of the 20th century. It was initially popularised by Ronald Fisher, and later by Pearson and Neyman who collaborated on a modified formulation. There are important philosophical differences between the two approaches, even though the mathematical methods underlying them are often identical (Lehmann 1993). Although personally I find Fisher's formulation of NHST more consistent, here I attempt to describe Neyman and Pearson's approach as I find it easier to discuss in relation to my experiments and I feel that an in-depth discussion of the entire controversy is not called for here.

In Neyman and Pearson's formulation of NHST, the first stage is to form a 'null hypothesis' and an 'alternate hypothesis'. In the prototypic case, the null hypothesis is that the 'treatments' or 'conditions' under investigation have no effect. The 'alternate hypothesis' is simply that they are not equivalent. After performing an experiment, the resulting data are analysed and a 'p-value' is calculated. This represents the probability of the data having been obtained, assuming that the null-hypothesis is true. If this value is very low (i.e. unlikely) then it is reasonable to reject the null-hypothesis and accept the alternate hypothesis. The difference between the conditions may also be said to be 'statistically significant'. Neyman and Pearson initially argued that prior to starting an experiment, the investigator should choose a 'significance level' α, and simply report whether or not the p-value was lower than this level (Lehmann 1993).

Errors in significance testing

Neyman and Pearson identified two main categories of errors that may occur: type I and type II statistical errors.

Type I statistical error

A type I error is a 'false positive' that occurs when the null hypothesis is falsely rejected – in other words if we state that there is a significant difference when in fact none should have been found. This type of error can occur by chance, and indeed probably does so relatively often given that even within a single experimental paper many individual p-values may have been calculated.

Type II statistical error

A type II error is a 'false negative' that occurs when we accept the null hypothesis – when in fact a significant difference does exist.

Arguably, in most cases where two values are being compared, they would be found to be statistically significantly different if the experiment was large enough, and well controlled enough to measure them with arbitrary precision. A failure to find a statistically significant difference therefore highlights a limitation of the experiment, rather than showing that no difference in fact exists.

For this reason, it is sometimes possible to carry out a ‘power’ calculation prior to commencing an experiment in order to demonstrate whether a given experiment would be likely to find a statistically significant difference between two or more parameters that are measured. For example, based on the results of a power calculation, an experimenter might decide to increase the number of samples or trials in an experiment, to be confident of finding a statistically significant p-value for any reasonably large effect. This obviously depends on the experimenter being able to predefine the size of a ‘reasonably large effect’, a judgement that in most circumstances will involve a degree of subjectivity.

Type III statistical error

Unlike other areas of mathematics, where it is often possible to obtain the same answer via numerous different methods, statistical methods are rarely ‘exact’ and different techniques, applied to the same data, generally return different results.

Florence Nightingale David proposed the concept of a third type of statistical error (David 1947). She points out that the correct statistical test must be justified and chosen prior to carrying out the analysis, otherwise the researcher is open to the accusation that she has performed many tests in order to choose the one that interprets the data most favourably!

A type III statistical error is not the same as performing an incorrect test. It could be that two tests are theoretically equally suitable, but always give different results from one another. If I perform both tests then only display the results from the one that interprets my data most favourably, this will introduce a bias in favour of my conclusions.

Calculating p-values – statistical methods

Any statistics textbook will describe a dizzying range of statistical methods, each suited to particular types of data. In this thesis, I use two psychophysical procedures: the method of constant stimuli, and adjustment matches. These are very different in terms of the type of data that they return, and must therefore be analysed by different methods.

Analysing data obtained by adjustment methods

The method of adjustment returns continuous data, each being a discrete continuous measurement of the sort that the family of ANOVA methods are well-suited to analyse. However, a number of assumptions must be met before ANOVA may be considered valid for analysing a given dataset. If these assumptions are not met, 'non-parametric' methods such as the Kruskal-Wallis test may be more suitable for the data. Given that I have only used standard ANOVA to analyse continuous data presented within this thesis, these alternative techniques are not discussed. However, I do describe the assumption-testing I have carried out prior to performing my ANOVA analyses.

ANOVA methods

The term 'ANOVA' covers a variety of methods. Variants of ANOVA include 'unbalanced ANOVA' which is able to analyse datasets where there are different numbers of datapoints for each group, and factorial ANOVA, which is able to return more complex analyses. For example, if I compare the performance of ten observers for two conditions, 2-way ANOVA can determine both whether the results differ significantly between observers, and whether the results differ significantly between groups, on a per-observer basis. 2-way ANOVA analyses also return an interaction term that may indicate whether the pattern of data differs between observers, or whether they all exhibit the same pattern of data.

As ANOVA is only used for a single set of experiments, reported in experimental chapter 3, the rationale behind the choice of ANOVA method utilized is described in this chapter, where it can be more readily related to the experimental design.

Interpreting interactions in ANOVA

My personal belief, and what I believe to have been the tradition throughout much of psychophysics, is that we generally repeat experiments on multiple individuals not because we cannot obtain sufficient data from a single observer for the purposes of analysis, but because we wish to ensure that the results obtained for our observer are representative of the wider population[1]. Pooling data, as ANOVA does, is only valid when we are reasonably certain that all the data show roughly the same pattern. ANOVA's usual safeguard against this, is to provide an interaction term. If the interaction term is small, then it seems reasonably likely that all the data are showing the same quantitative pattern. However, are interaction effects an appropriate tool for testing weak predictions that are expressed in the form of inequalities, and where one may expect to find quantitative but not qualitative inter-observer variation?

It is relatively easy to generate artificial datasets that suggest that interactions are not reliable indicators of heterogeneity within a population. For example in a population of nine observers who show no effect (condition A = condition B) and one who shows a substantial effect (condition A < condition B), the atypical individual may skew the results just enough that the whole population may show an effect, when most individuals show none, even though no interaction effect is found. Conversely, ANOVA may fail to show either a main effect or a statistically significant interaction, even when the effect in the atypical observer is readily apparent, whether the data is eyeballed or whether a T-test is performed and a Bonferroni correction is applied. In preliminary simulations, I obtained such misleading results on up to 20% of occasions and although I do not report further details here, numerous examples of such apparent paradoxes may be found in experiment 3 where both individualised and group analyses are reported.

[1] For a deeper discussion of this position, see page 98.

Findings from inferential statistics are by nature probabilistic and therefore sometimes incorrect. Is it unfair to caution against the use of ANOVA simply because artificial datasets can be constructed where the method frequently fails? And are there any alternative techniques that might be used? Beauchaine surveys techniques that can be used to detect multiple classes of individuals within a population (Beauchaine 2007); however many of these techniques require extremely large numbers of individuals. Psychophysics experiments may be likened to meta-analyses used in biomedical research to combine data from separate studies, in that the findings for one observer are conceptually independent of those obtained for others – much as there is no paradox if two studies in a meta-analysis show different results because they were carried out on two different populations. However, such heterogeneity tests have been criticised as being generally misleading. For small samples, they are not sufficiently sensitive to detect large degrees of heterogeneity; for large samples statistically significant degrees of heterogeneity will nearly always be detected (Takkouche, Cadarso-Suarez et al. 1999).

It is also my view that ANOVA interaction terms are poorly matched to the testing of hypotheses that are not quantitative in nature. For example, imagine that we were to test the hypothesis that children always grow at a positive rate (i.e. they do not shrink). If we measure some children at two points in time, our general knowledge tells us that some will have grown much taller due to a growth spurt, whereas others will hardly have grown at all. The interaction term is therefore likely to be large, however this does not negate our likely finding that children grow taller with time, and it tells us nothing about whether or not there were some individuals who actually shrank. In other words, ANOVA is testing a different hypothesis from that which was specified in advance[2]. Furthermore, even if we correctly dismiss our finding of a statistically significant interaction as meaningless (with respect to our hypothesis), our conclusion that children can only grow taller will remain formally untested and is as subjective as it was prior to performing the statistical analysis.

[2] On the other hand, if the interaction term was very small this could be argued to suggest that there were no qualitative differences between the way in which the children grew, as an absence of quantitative differences implies a lack of qualitative differences. However, it should be remembered that significance tests cannot prove equality (i.e. even if the interaction term was small, this does not prove there is no difference). My overall point is that although it is arguably not useless to calculate this interaction term, its interpretation is not straightforward.

Without good alternatives it seems important to continue to cautiously interpret interaction effects and also to eyeball the raw data. I have also reported individualised analyses alongside the overall results for some sets of experiments.

Assumption testing in ANOVA

To use standard, parametric ANOVA, the data should satisfy a number of criteria pertaining to their distribution. The two main criteria are that they are normally distributed, and that each group has equal variance. Many statistical tests have been proposed to determine whether these assumptions are true. Their use is not universally accepted – Box and Anderson point out that when assumption tests fail to demonstrate violations of assumptions, this may often be due to a lack of statistical power (Box and Andersen 1955). Assumption testing cannot show that our assumptions are met – only that they are not. Daniels therefore counters that when assumptions are grossly violated, it is as well to know about it, and assumption testing should generally be performed (Daniels 1955). Nonetheless, I have run assumption-testing algorithms independently for each condition examined using ANOVA, and the results for each test are included in the tables of data reported for each sub-experiment.

Testing the assumption of the normal distribution

The first and probably the most significant assumption is that the data are drawn from a normal distribution. There is considerable controversy both about how best to assess normality, and at least an equal amount about how to use these assessments once they have been performed. If the data is clearly not normally distributed, then it would be better to use a nonparametric method such as the Kruskal-Wallis test; the problem is that the tests used to assess normality are very sensitive, and can often indicate with a high degree of statistical significance that a dataset is not normally distributed, even when its deviation from normality is so small that it would be perfectly valid to analyse it using a parametric method.

Normal distribution tests are based on mathematical tests of its third and fourth moments around the mean – namely variance, skewness and kurtosis. Normality is accepted if the values estimated for these parameters lie within given limits. Cutoffs that have been suggested, are ‘’+/- 2’ for skewness and ‘1 to 5’ for kurtosis (Peat and Barton 2005).

More sophisticated methods of assessing normality such as Z-scores are often calculated using these variables (McQueen and Knussen 2006), so I have reported them individually here for each condition for each observer. There are several conventions on how to report kurtosis. The kurtosis of a normal distribution curve is '3', but it is common to normalise it to zero by subtracting 3. The Matlab function for kurtosis does not do this, and the values I have reported here should tend to '3' for normally distributed data.

Homogeneity of Variance

When comparing two or more conditions it is necessary to check whether they have equal variances – that is, whether they are homoscedastic. For the relevant datasets, I have made two sets of comparisons to match the analyses that I plan to perform: one on an individual basis for each observer, and one comparing both observers and conditions. For individual observers I report p-values for homoscedasticity, as determined by Levene's test (Levene 1960). If they are <0.05 then the requirement of homoscedasticity for that observer has not been met, at that level of significance.

Perhaps the two most commonly used tests of homoscedasticity are Levene's test and Bartlett's test. Bartlett's test is both more powerful, and more vulnerable to violations of the assumption of normality (Lim and Loh 1996). I chose Levene's test simply on the basis of ready availability of a computer program that could compute it (Trujillo-Ortiz and Hernandez-Walls 2003).

For the most part, individual observers have similar variances for each condition. However, this is not true for comparisons between observers – perhaps unsurprisingly as amongst other factors the amount of variance in an observer's settings is probably related to the amount of experience they have performing the experimental task, and may also be modified by their judgments about what constitutes an acceptable match.

The question then is: having shown that for the inter-observer comparisons the requirement of homoscedasticity is not met, should this modify the analysis that we perform? A problem with the test, is that it is able to reliably detect violations of homoscedasticity at levels that are far too small to significantly impact the validity of ANOVA performed on the data. For this reason, it is desirable to examine the confidence intervals independently, and only revert to a non-parametric test if they are 'different by orders of magnitude'. Standard deviations are reported individually for each observer for each condition, and a cursory examination suggests that for the most part, they are actually relatively close, differing at most by a factor of 2-3 between observers.

It is worth bearing in mind that violations of homoscedasticity are likely to cause a type I error (false positive) rather than a type II error (false negative) (Box and Andersen 1955). The majority of the results in the third experimental chapter are negative so violations of this assumption have probably had little bearing on our conclusions.

Analysing data obtained by the 'method of constants'

The 'method of constant stimuli' returns categorical data, and in my experiments these take the form of binary values that are fit by a psychometric curve, as previously described in the 'methods' section. Method of constants data are particularly difficult to analyse using inferential statistics and until recently, statistical analyses do not commonly appear to have been performed on experiments of this type – for a brief survey of the use of inferential statistics in visual psychophysics see Appendix D.

After searching through many papers and books on statistics and seeking a consultation with a statistician, I was able to identify only four methods that might be used to indicate statistical significance for 'method-of-constants' data. Their relative merits are discussed at length in Appendix E. Of the four methods, two appeared largely or completely invalid, and the other two have significant limitations in comparison with standard ANOVA.

The most intuitive method of performing statistical analyses is perhaps simply to compare error-bars on graphs. However, a Monte-Carlo method 'pfcmp' is available that is able to provide p-values for selected pairwise comparisons. As such, these methods are more comparable to a T-test than to ANOVA as neither is readily expandable in order to compare more than two conditions, or to analyse results from multiple observers. This is not necessarily a drawback, however. My personal view, discussed later in this chapter, is that it is frequently undesirable to combine data from multiple observers.

Criticisms of NHST

Although significance testing is now almost ubiquitous, an over-reliance on it has been criticised in on many grounds. Often it is argued that authors simply misunderstand the techniques they are using: even statistics lecturers often entertain misconceptions about the methods they teach (Haller and Kraus 2002). Other authorities go further, arguing not merely that significance testing needs to be better understood by its practitioners, but that it is philosophically flawed and should probably be used rarely if at all (Carver 1978; Loftus 1991; Cohen 1994; Thompson 2001; Gigerenzer 2004).

This controversy is perhaps particularly keen in the field of psychology. In his book review, 'On the Tyranny of Hypothesis Testing in the Social Sciences', Geoffrey Loftus coined the term 'miscast objectivity', arguing that hypothesis-testing "provides the illusion of endowing psychological data – which are intrinsically complicated, messy, multidimensional, and subjective – with a deceptive simplicity and objectivity" (Loftus 1991). In the following discussion I concentrate on several issues that seem particularly pertinent to psychophysics methods.

The fallacy of 'insignificant significance'

The 'fallacy of insignificant significance' is a term frequently used in medical statistics, often to refer to the circumstance when a drug or intervention can be shown to be more effective than a standard treatment, but when on closer examination the difference demonstrated is real but of little practical consequence (Skrabanek and McCormick 1998).

This fallacy has its counterpart in the field of psychophysics, where the ability to collect large amounts of data makes it possible to measure and study vanishingly small differences (Cohen 1988). This has caused some researchers to be sceptical as to the value of a great deal of psychology research, and has led several publications to require the reporting of effect sizes. Since 1994, the American Psychological Association has recommended that effect sizes be reported in psychological literature (APA 1994). A few journals *require* that effect sizes be reported in addition to p-values stating whether or not the differences being studied were statistically significant. For reviews, see Thompson (Thompson 1998) and Loftus (Loftus 1991).

The near-certain falsity of the null-hypothesis

In the field of soft-psychology, Meehl and Lykken have described what Lykken called 'ambient correlation noise' (Cribbie, Gruman et al. 2004) and Meehl called the 'psychological crud factor': "The crud factor consists of the objectively real causal connections and resulting statistical correlations that we would know with numerical precision if we always had large enough samples." (Meehl 1997). This causes almost *anything* that is investigated, to prove statistically significant if sufficient data are collected.

If there is a psychological crud-factor, then might there also be a psychophysical crud-factor, and what form would it take? The exact answer would depend on the form of the experiment, but in essence it consists of all those biases that we try, never entirely successfully, to eliminate. If I introduce some new element onto a computer display, this will inevitably change the load on the monitor's power supply and cause the overall picture to change brightness. Light reflected inside the monitor will slightly change the colour of every area of the screen. As an observer learns to perform an experiment, his threshold will almost certainly change, and if all conditions cannot be included within the same blocks of trials, a selective bias will result. All these factors will inevitably conspire to change the observer's measured thresholds – at least slightly. If sufficient data are collected, then these changes will also cause the result to become statistically significant, which if it happens consistently may be viewed as direct detection of the bias, or if less consistently as an increase in the type-I error rate.

I would be willing to accept that, provided all equipment is well calibrated, psychophysical bias is liable to be much smaller than that present in other branches of psychology. But in contrast, it is easier to collect psychophysical data than many other varieties. It is a major undertaking, for example, to find 20 individuals with the same variety of schizophrenia and administer a questionnaire to each of them. Yet in a psychophysics experiment a single observer may perform a thousand trials in an hour.

It is often pointed out that it is perverse to attempt to disprove a null-hypothesis which is almost always false. But the most serious issue is that if the null-hypothesis is never true, then the more data is collected the more likely we are to obtain a statistically-significant finding. If we accept this, then we must admit that all statistical errors are 'type II errors'. If we already know that the null-hypothesis is false, then why do we test it? Meehl framed this as a paradox – that the more data we collect, the less meaningful a 'significant' result becomes (Meehl 1967).

The problem of demonstrating equivalence

One of the potential patterns my data might show is for there to be no difference between the control and test conditions. This should not be interpreted as a negative result – in scientific terms negative and positive findings can be equally important. However, most methods of significance testing are only able to demonstrate significance, not equivalence.

P-values are related to the probability of a positive result occurring by chance, but the common misconception that they are the inverse of the probability of the null hypothesis being true is incorrect. In claiming that there is no significant difference between the control and test conditions we incur a 'reversal of the usual scientific burden of proof' (Dayton 1998).

A common approach is to perform post-hoc power analysis calculations to determine our expectation of obtaining a positive result of a prespecified size. Although this approach is widely used, Hoenig and Heisey show that it is untenable (Hoenig and Heisey 2001). They also address the question of whether it is possible to determine the 'detectable effect size' – the smallest effect that might be detected with a specified power (e.g. 0.95). Again, they conclude that the methods used to calculate it have not been formally justified and that the approaches currently used are severely flawed.

One approach that has found acceptance is to use tests of equivalence that depend on a predefined ‘significant interval’ being set, in much the same way as for power analysis. One of the more commonly used approaches is described by Schuirmann (1987), and consists of performing two one-sided t-tests on the data. Berger and Hsu give a discussion of several alternative equivalence tests, decrying their lack of power, before proposing a further test (Berger and Hsu 1996). They also mention the difficulties of performing multiple-comparisons when using these methods.

As applied to my data, equivalence testing has several pitfalls. As with power testing, the problem arises of how to define a significant difference. The only obvious measure would be to employ a single just-noticeable-difference. Yet I believe that this would be too small – smaller even than artifacts due to monitor calibration and observer drift. Seemingly perversely, it is also quite possible for a test of equivalence and a null hypothesis test to return an identical result, whether positive or negative, when applied to the same data (Cribbie, Gruman et al. 2004).

Such a result would not be paradoxical. In the case where both results are negative, this merely shows that the tests have insufficient power to draw any conclusions. If both results are positive, then it is likely that the data groups are statistically significantly different, but that the difference is smaller than the threshold previously chosen as indicating practical significance. Given that this decision is generally arbitrary, this provides less information than a careful examination of confidence intervals, combined with a positive significance test. This approach also seems more satisfactory to me than objectively testing whether the data exceed a subjectively chosen threshold. In view of this, I feel that equivalence testing is redundant in my experiments, and I have chosen not to attempt any equivalence tests for data presented here.

Practical science and the validity of statistical tests

In Popper’s book, ‘The Logic of Scientific Discovery’, he outlines the hypothetico-deductive method of scientific enquiry thus: ‘the theoretician puts certain definite questions to the experimenter, and the latter, by his experiments, tries to elicit a decisive answer to these questions...’ But Asimov – who was a scientist and science writer as well as a novelist – reputedly put forward what is to my mind a far more realistic view of laboratory science: “The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!', but 'hmm… that's funny ...’”.

Popper admits: "Every experimental physicist recognises those surprising and inexplicable apparent 'effects' which in his laboratory can sometimes even be reproduced for some time, but which finally disappear without trace". But these effects do not always disappear, and there are many important areas in which observation has preceded theory. A classic example would be the mass discrepancy observed by Aston when a heavy atom is compared to lighter isotopes with the same net number of neutrons and protons (Rhodes 1986). This observation ultimately proved vital to nuclear research (p. 260). Perhaps – according to the hypothetico-deductive formulation – this work did not constitute 'science'. Yet it was a vital part of the investigative process.

Many areas of research – for example epidemiology and drugs trials – require meticulous long-term planning, and ask narrowly defined and carefully chosen research questions – conforming perhaps more closely to the hypothetico-deductive view of scientific progress. But in the laboratory it is often the case that an individual experimenter acts as both technician and theoretician, recursively searching for substantial effects. Where necessary, one may create original theories to explain unexpected findings, then perform a new or modified experiment in an attempt to build up a fuller picture of the phenomenon under investigation. Ultimately, when a coherent theory has been developed that is capable of making testable predictions, Popper's notion of falsifiability becomes important[3].

This alternate view of how science should progress has many advocates. Don Jewett's 'strong-inference plus' divides research into an exploratory phase, a pilot phase, and a hypothesis testing phase (Jewett 2005). Ioannidis refers to a 'hypothesis generating phase' of experimentation (Ioannidis 2005). Chamberlin decries the investigation of single hypotheses, advocating that researchers should entertain as many plausible hypotheses as possible, in order to avoid premature emotional commitment to incorrect theories (Chamberlin 1965). I regard my own research as being closer to exploration or a pilot phase than a hypothesis testing phase. Some of my experiments are capable of producing data patterns that do not correspond to any of my tentative predictions. Does formal hypothesis testing make sense under these circumstances?

[3] Other formulations of the hypothetico-deductive model also allow for confirmation.

Textbooks of statistical hypothesis testing emphasize how important it is to correct for the number of tests performed, and how it is imperative to decide in advance which comparisons should be made. But even if this is done, if the results of a large number of speculative pilot experiments are simply discarded or filed for later reference, then it surely becomes impossible to calculate the corrections that should be made to the significance tests.

This thesis contains the results of perhaps one-half of the larger-scale experiments that I performed. In addition I carried out a large number of pilot experiments using only myself as an observer, in an effort to discover which ideas were worth pursuing. This leaves me open to the accusation that I may have chosen to report or pursue only those ideas that confirmed my own beliefs and prejudices. I must plead guilty, yet I suspect the accusation could be equally fairly leveled at most other people in the field. And had I acted otherwise, it might arguably have only served to pollute an already confused literature with uninteresting, marginally statistically significant experiments.

In his paper 'Why most published research findings are false', Ioannidis discusses these issues, demonstrating mathematically that "The smaller the studies conducted in a scientific field, the less likely the research findings are to be true." And that "The greater the number and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true" (Ioannidis 2005). Yet he recognises that it is impossible for 'hypothesis generating experiments' to be large-scale.

Related issues are also well known within the fields of drug-testing and parapsychology where 'publication bias' and the related 'file-drawer problem' (Rosenthal 1979) are known to be a serious cause of error whenever literature-reviews are performed or meta-analyses are published. Unfortunately the alternative – to slavishly follow a regimented and meticulously planned series of experiments – would leave little room to investigate those serendipitous findings or speculative pilot experiments which have historically proved so fruitful.

I am not arguing for a lack of method or rigor; that no consideration should be accorded to experimental noise or that experiments should be performed in an haphazard, thoughtless manner. What I am arguing is that in the exploratory and pilot contexts in which I carried out my experiments, strict interpretations of statistical metrics such as p-values do not make a great deal of sense. The problem of how to correct for these unquantifiable but undoubtedly real factors is simply intractable.

Ultimately, if hypothesis tests are to be employed, I see this issue as a strong argument for Fisher's practice of reporting absolute p-values (Fisher 1955), rather than simply citing 'significant' or 'not-significant' as is done in Neyman-Pearson's formulation. This maximises the amount of information available to the experimenter when deciding how much weight to accord a result and reminds the interpreter of the role that subjective judgment plays in science, rather than downplaying it inappropriately by allowing the decision to accept or reject a finding to be made 'by procedure'. Secondly, these sources of uncertainty lead me to be extremely skeptical of results that fail to attain statistical significance by a much better margin better than the commonly used 5% level.

The problem of imprecise hypotheses and weak predictions

Popper describes how theories may be described as weak or strong, depending on the degree of jeopardy they are placed in by an experiment (Popper 2007). Given the near ubiquitous falsity of the null-hypothesis, a weak qualitative prediction such as stating that induction will be decreased in some given situation has a 50% chance of being either confirmed or disconfirmed by a sufficiently powerful experiment (Meehl 1967). A more precise prediction – for example that induction will be decreased by 43% +/- 0.1% – has a much greater chance of being disproved. An experiment confirming this prediction would represent a powerful test of a strong theory. Clearly it is desirable to make predictions that are as precise as any given theory will allow.

In its early stages, psychological science is rarely able to make predictions that are as precise as we might wish, and this is the case for the majority of my theories. Also, as previously discussed it is a fallacy to suggest that only hypothesis-testing research is of merit. Take a prediction that differential motion should reduce induction. Assuming the margin of error is small, a consistent demonstration that motion reduces induction by 10% should surely be recognised as being very different from a result that motion reduces induction by 100%, even though the same weak prediction was made in both cases.

Hypothesis testing is typically employed to determine whether statistical differences are 'significant' or 'non-significant'. However, methods also exist whereby p-values may be calculated for 'imprecise hypotheses' (Pitz 1978). Unfortunately these methods are fairly complex and unwieldy. In my view they are simply too difficult to use correctly, considering the problems that are already evident in applying significance tests to method-of-constants data. I am also skeptical about the utility of applying precise tests to subjectively chosen thresholds. Ultimately, for complex experiments such as my own I do not believe that these techniques are viable.

Combining data from two or more observers

Classical statistical techniques generally pool data from many observers. This is because there are usually limits on how much data can be collected from individuals, so it is necessary to have many subjects in an experiment. For example, in a pharmaceutical trial one cannot generally cure an individual patient, make him ill again, then retest the drug. An approach similar to this, termed 'crossover trials' can sometimes be performed, but it is rare to be able to repeat an experiment hundreds of times on the same individual, as is our privilege in psychophysics.

As a consequence, psychophysicists may frequently make statements that are almost certainly true for individual observers... e.g. ‘Observer A is a trichromat’, ‘Observer B is a dichromat’ and ‘Observer C is an anomalous trichromat’. Our findings for each observer should generally be considered independent, and the reason we include multiple observers in an experiment should normally be to establish the generalisability of our findings throughout the population rather than to attain statistical significance. It would be ludicrous to state that ‘on average, observers A, B and C are anomalous trichromats’. Combining data from several observers for analysis is only likely to be valid if a number of assumptions are met. Firstly, that we have no interest in characterising or noting any inter-individual variations. And secondly, that we already have strong grounds to expect all of the observers to show broadly the same pattern of results.

It is often the case that “pathology illuminates normal physiology”, so any missed opportunity to identify alternative patterns is to be regretted – even if they are normal variants and not disadvantageous in any way. Perhaps more importantly, if qualitative inter-individual differences are present this might severely distort the results of pooled data-analyses – see, for example, ‘Simpson’s paradox’ (Malinas and Bigelow 2004). Thirdly, it is my view that asking additional observers to perform an experiment may be considered a weak form of experimental replication (Lindsay and Ehrenberg 1993): it is therefore far more persuasive to show that *n* observers each exhibit a particular phenomenon than it is to show that when their data are pooled, on average *n* observers exhibit a particular phenomenon.

If it is only valid to combine data from several observers for analysis when it is reasonably certain that most observers will show qualitatively the same pattern of results, how certain can we be that this is the case for the experiments reported here? The answer may depend in part on the level of visual processing that we believe we are studying. Intuitively one might expect high-level processes to be more likely to vary between observers than more fundamental processes, and I would argue that the processes investigated here such as binocular disparity and motion-based segmentation are likely to be relatively low-level phenomena. On the other hand, even at the lowest level of visual perception: that of the photoreceptor, qualitative inter-individual variations (e.g. anomalous colour vision) are reasonably common. I would also point to the data from my first experiment, which does show qualitative inter-individual differences.

The uncomfortable situation is that if there is sufficient statistical power to determine whether individual observers show the same pattern, then there is little purpose in performing a pooled analysis. Conversely, if there is insufficient power to analyse individualised data, one must assume that that each observer shows the same pattern of results in order to justify performing a pooled ANOVA. In the context of evaluating an as-yet-unstudied relationship between two phenomena, the validity of this assumption is open to conjecture. In any case, if data from *n* observers were required to be pooled before the data became statistically significant, it would be reasonable to ask why a single observer was not asked to make *n* times as many observations in order to avoid forcing the assumption of observer homogeneity.

Although I cannot find any explicit discussion of these issues in the literature, it seems to me that the approach of analysing data on an individual basis is implicitly supported by the way most experimenters have conducted and reported their findings up until now. Psychophysics data are often reported on an individualised basis, frequently for relatively small numbers of observers (Webster and Mollon 1991; Zaidi, Yoshimi et al. 1991; Zaidi, Yoshimi et al. 1992; Chichilnisky and Wandell 1995; Werner, Sharpe et al. 2000), or even individuals (Valberg 1974). Combining data from potentially heterogeneous individuals may be a necessary evil in many branches of science, but it is not one that psychophysicists are always obliged to indulge in.

Adjuncts and alternatives to NHST

Confidence intervals and error bars

Confidence intervals, already discussed in relation to significance testing, provide visual estimates of the error attached to a measurement. According to a frequentist interpretation, on 95% of occasions that a confidence interval is drawn, the true mean will lie within its bounds. Error bars may also be drawn on graphs to show the standard error of the mean, or the standard error, or 50% or 99% confidence intervals, and it is clearly important to know what they represent before interpreting them.

Confidence intervals are often said to be easier to interpret than p-values, and those authors who decry the use of NHST often recommend them as an alternative (Loftus 1991; Cohen 1994; Meehl 1997; Hoenig and Heisey 2001). I have not managed to find any evidence that they are better understood, and the fact that there are widespread misconceptions about their definition or utility in determining statistical significance suggests that this view may be misplaced (Smith 1997; Ludbrook 2000; Belia, Fidler et al. 2005).

A metrological approach to data analysis

Take an encyclopedia and look up the heights of the ten highest mountains, and you are likely to find a neatly ordered list, devoid of confidence-intervals, p-values, multiple-comparison corrections and discussions of whether the multiple-comparison correction should be made only for the 10 mountains shown, or the hundreds of candidate peaks that might have proved to number amongst the top-10, or even the millions of 100-foot hills that are manifestly not amongst the tallest mountains in the world – but how do we know without measuring them? Yet a textbook of metrology describes how any measurement can be expressed in terms of a fixed term, a Gaussian error term and a bias term (Allan 1997). Although one may think of a surveyor's measurements as being unitary, they are almost invariably averages of repeat readings. In many ways this is strongly reminiscent of the situation in psychophysics when we measure a threshold through repeated sampling.

It would be perfectly feasible to perform significance tests to determine how certain we are that Mount Everest (8850 metres) is taller than K2 (8611 metres) but in practice we do not, because we know that the error terms are very small relative to the size differences between the peaks. Any challenges to the rankings of these two mountains are more likely to be the results of mistakes than statistical uncertainty. When K2 was briefly claimed to be higher than Everest, this was the result of some batteries in a satellite-altimeter going flat (Lak 2000), which is exactly the kind of flaw that a significance-test would be unable to deal with.

Why should it sometimes be possible to state that p-values are significant, without formally calculating them? The answer is that we can estimate our measurement error, in the form of standard-errors-of-the-mean or 95% confidence intervals. As discussed in Appendix E (pg. 349), when two confidence-intervals are just touching, the results are already likely to be highly statistically significantly different. As the two population means are drawn further apart the p-values, if calculated, would drop rapidly. Table 3 shows the proportion of values from a normal distribution that lie within 4 standard-deviations of the mean. Recalling that 95% confidence intervals are normally distributed, and that two populations are statistically significantly different at the 95% level when their estimated means are separated by only 1.39 standard errors of the mean (which is a normally distributed variable, as dictated by the law of large numbers), it is clear that by the time two thresholds are separated by three or more standard-errors of the mean, they are likely to be statistically significantly different to a very high degree.

Number of standard deviations *n*	Probability of the true mean lying within n deviations of the estimated mean	Probability of the true mean lying without n deviations of the estimated mean
1	0.683	0.317
2	0.975	0.025
3	0.997	0.003
4	0.999937	0.000063

Table 3: The proportion of the area of a normal distribution curve that falls within n standard-deviations of the mean.

In contrast, the nature of multiple-comparison corrections for p-values is that they are at worst directly proportional to their number (e.g. the Bonferroni correction (Lindman 1992)), so provided that the putative multiple-comparisons seem reasonably few in number and that the means in question are separated by a reasonable number of standard-errors, I believe it is reasonable to accept that they differ significantly without performing further calculations.

It is gratifying to calculate infinitesimal p-values when the format of the data is such that it is effortless to do so. However, science is a human endeavour and when putting a numerical value on the validity of a hypothesis is difficult and time consuming and the risk of auxiliary hypotheses being false ('the altimeter batteries are fully charged', 'my computer monitor's calibration is flawless') is likely to greatly outweigh the probability of the significant p-value being due to chance, then I argue that it is actually wrong to attempt to do so. The time spent in this undertaking would be better allocated to characterising calibration flaws or performing additional control experiments. Equally importantly, a miniscule p-value may be taken to emphasize the validity of the finding, drawing attention away from the design and implementation of the experiment – which is where the majority of the scrutiny should be directed.

A similar argument is made by Ord et al. who show that, when calculating the risks of rare events, errors in logic and calculation are much more significant contributors to uncertainty than the numerical estimates of the calculation error (Ord, Hillerbrand et al. 2009). They agree that addressing the problem using multiple methods (akin to performing a highly 'differentiated replication' (Lindsay and Ehrenberg 1993)) is the best way to catch these errors.

Effect size - Cohen's 'd' statistic

Cohen's 'd' statistic is calculated by dividing the absolute difference between two means by their common standard-deviation (Cohen 1988). The resulting statistic is dimensionless and may therefore be used to provide a 'portable' measure of the magnitude of any difference between two means. Cohen points out that differences that might be considered highly significant in one field, may be inconsequential in another. However, as a rough guide he suggests that 'd' values of 0.2 would generally be considered 'small', 0.5 would be 'medium' and 0.8, 'large' (Cohen 1988).

However, the major drawback of Cohen's effect-size metric is that it is not possible to obtain estimates of the standard-deviation from pfit, so it is impossible to calculate 'd' statistics independently for experiments using the method of constants. Analysing the standard-deviations between sessions would suffer from even worse drawbacks than when this method is employed with ANOVA, in that the value obtained would be entirely dependent on the number of trials per session and would eventually approach infinity with an increasing numbers of trials per session.

The obvious conceptual equivalents would be to either to compare the difference between two 50% thresholds to the mean slope of the two psychometric functions, or with a just-noticeable-difference in that part of the stimulus-space. However, the JND cannot be extracted from the original data, and measuring it would therefore involve a significant amount of additional work. Using either of these methods in turn precludes comparing these 'd' statistics directly with those obtained using the methods of adjustment, as the 'small', 'medium' and 'large' cutoffs would in all probability be considerably different. In this case, there seems no advantage over simply calculating the raw effect sizes.

A ‘drawback’ of Cohen’s ‘d’ statistic may provide a workable solution. It has been criticised as being partly dependent on the variance of the method of measurement employed. Although Cohen points out that significance-tests such as ANOVA are also influenced by measurement error, this is still a serious issue. For example, in my third set of experiments where it is straightforward to calculate the ‘d’ statistic, the standard-deviation of the moving-target condition is considerably smaller than the standard-deviation of the moving background condition. I suspect this is probably due to the difficulty of the task, but it still causes the ‘d’ statistic calculated to be different in each case even though there is no obvious reason why a given perceptual difference should be more or less practically significant in either condition. However, provided the reader only looks at the approximate magnitude of the effect size rather than its absolute value, it may still represent a useful metric. A positive example of its utility can be illustrated by examining the ‘d’ statistic for observer YL’s data in experiment 3, moving target condition. Although she is less subject to simultaneous contrast induction than most other observers and the differences between her thresholds for different conditions are smaller, she actually obtains a similar value of ‘d’ because her discrimination appears to be superior and the standard-deviations of her settings are lower.

It is possible to ‘normalise’ the y-axis (effect size) for most of my experiments, so that a ‘1’ represents complete induction and a ‘0’ represents zero-induction. In these circumstances it is also possible to compare the absolute magnitude of any change in induction directly between both the 2afc and adjustment experiments and thereby make direct comparisons of the effect sizes within these series of experiments.

Given that even when calculated in the standard manner, the ‘d’ statistic is often strongly dependent on the experimental task, it does not seem wholly unreasonable to estimate a ‘d’ statistic for method-of-constants experiments using the standard-deviations obtained in an adjustment experiment and the 50% thresholds determined using the curve-fitting procedure. This does depend on two assumptions: firstly that most of the variance in the data reflects intrinsic uncertainties in our visual processing, rather than extrinsic sources such as quantisation noise in the adjustment dials or limitations in motor control that are significant in ‘methods of adjustment’ but not in ‘methods of constant-stimuli’. And secondly, that the degree of intrinsic ‘noise’ in our visual systems is similar in both experiments. Neither of these assumptions is likely to be absolutely true, but as I used broadly similar stimulus sizes, colours and brightnesses in all my experiments I think it can be argued that it is reasonable to calculate the statistic using this method, given the potential wide variability of the standard deviation even when used ‘by the book’.

Statistical techniques that are undoubtedly well-suited to other fields often appear cumbersome when applied to psychophysics, and this appears equally applicable to Cohen’s effect-size metric ‘d’. However, whilst cognisant of its disadvantages and anxious that the figure not be over-interpreted, I have chosen to calculate and report it whenever possible. In general, this means that it may only be calculated for observer’s ACH, KW and YL, as we were the only observers to take part in most series of experiments.

Replication

All authorities seem to agree that replication is a cornerstone of the scientific method. As Fisher stated, ‘no isolated experiment, however significant in itself, can suffice for the experimental demonstration of any natural phenomenon... In relation to the test of significance, we may say that a phenomenon is experimentally demonstrable when we know how to conduct an experiment which will rarely fail to give us a statistically significant result.’ (Fisher 1935). Cohen argues that, ‘Given the problems of statistical induction, we must finally rely, as have the older sciences, on replication.’ (Cohen 1994)

My personal view, previously stated, is that psychophysics experiments such as my own whose aims are to make qualitative statements about how the visual system works, can generally provide results that are independently valid for each individual observer. Whenever data are collected for an additional observer, this may be understood as a replication of the original experiment rather than as an extension of it.

Lindsay et al. discuss strong and weak varieties of replication (Lindsay and Ehrenberg 1993). Whilst it is ultimately desirable for replication to be performed independently by other experimenters, literature reviews in many scientific fields suggest that replications are not performed nearly as frequently as might be considered desirable (Lindsay and Ehrenberg 1993; Muma 1993). Aside from reporting individualised data, I report several experiments that ask related questions, but differ slightly in methodology or measurement technique. When applicable, this is discussed in the text.

Summary

On reflection, I have devoted what may seem a disproportionate amount of effort and discussion to the problem of applying significance tests to my experiments. This was never my intention, but reflects the outcome of a journey during which I was wrestling to understand the techniques I was using, in order to be certain that I was applying and interpreting them correctly.

My personal view is that I do not adopt the absolutist position against hypothesis testing – I am prepared to accept that it is valid when used in appropriate situations and interpreted correctly. However, I remain confused about whether to apply it to my own experiments, and I am doubtful that the p-values can be meaningfully interpreted after they have been calculated. I make no apology for this. Although I have specifically searched out and read articles supporting hypothesis testing, e.g. (Muliak, Raju et al. 1994; Knapp 1998; Levin 1998), I personally felt that the critical articles were generally more convincing. Several were written by doyens in the fields of statistics (Jacob Cohen) and psychology (Paul Meehl) so whilst the notion that hypothesis-testing is flawed may not be mainstream amongst experimentalists, it finds support amongst respected theoreticians. Yet the real impression I take away with me is one of surprise that the debate is so acrimonious, and that after nearly a century it is still unresolved. There is no shame in admitting that I remain confused about many points of the arguments.

Having performed most of the statistical tests in this thesis post-hoc, it is also interesting to look back and see whether they changed any of my conclusions:

Several results in the third set of experiments, which I had previously assumed to be statistically insignificant, turned out to be statistically significant. However, my judgment that the differences are practically insignificant is wholly subjective, and remains unchanged. These results are discussed in the text.

None of the results I had previously claimed to be ‘positive’ turned out to be statistically insignificant, and although this is reassuring it is not surprising. Given that the confidence intervals for effects I considered of practical importance were often separated by distances several times the size of the 95% confidence intervals, it was obvious that their p-values would be infinitesimal even if they were difficult to calculate.

Chapter 5, Experiment 1: Global Scene Attributes

Introduction

Both computational models of colour-constancy, and experimental investigations of the phenomena, show that the brain must integrate information from across the visual scene if it is to assign the correct colour to a target object to which it is attending. Yet classical studies of chromatic induction show that long-range effects are almost non-existent. Likewise, increasing the size of the background increases the influence of the background, but at an exponentially decreasing rate. Induction is largely complete when the background radius is just 1 degree larger than the test field. Conversely, when the inducing field is separated from the test field by even half a degree, induction is almost eliminated (Walraven 1973). These results could imply either that simultaneous chromatic contrast induction does not have a major role in colour constancy, or that colour constancy works by a mechanism very different to that previously assumed.

Wachtler et al. (2001) performed an experiment in which surfaces remote from the target modulated the strength of simultaneous chromatic induction. Furthermore, the absolute distance of these objects from the target seemed not to affect the magnitude of its influence, a hallmark of computational mechanisms of colour-constancy that should act or not depending on the validity of different surfaces as markers of the relevant illumination framework, and not on their retinal distance from each other. The paradigm employed by Wachtler et al. is probably most closely comparable to that used by Nascimento and Foster (2001), who used a nulling method in combination with the method of constant stimuli to study chromatic shifts induced by discrete, near-instantaneous illumination changes. Their stimuli added some remote surfaces to the background, then changed their surface chromaticities in a way that was either consistent with the illumination change, or not. Their result showed that induction was strongest when the remote surfaces changed in a way consistent with the background illumination change, and weaker when the remote surfaces chromaticities did not change. This difference is the key effect studied in their paper. However, it was only apparent when the background colour changed; there was no measurable effect on the colour of the target if the background colour remained constant, and only the remote surfaces changed colour in a way indicative of an illumination shift. This demonstrates that the effect was closely linked to chromatic induction by the local background.

We decided to design a set of experiments based on the paradigm used by Wachtler et al (2001). We wished to expand on his original experiments to test the prediction that increasing the number of distinct surfaces within a scene should increase the extent of simultaneous chromatic induction. This should theoretically improve their reliability as cues for colour constancy (Hurlbert 1998). Secondly, I wished to broadly determine which parts of the brain were involved in the effect, by investigating binocular interactions.

Experimental design

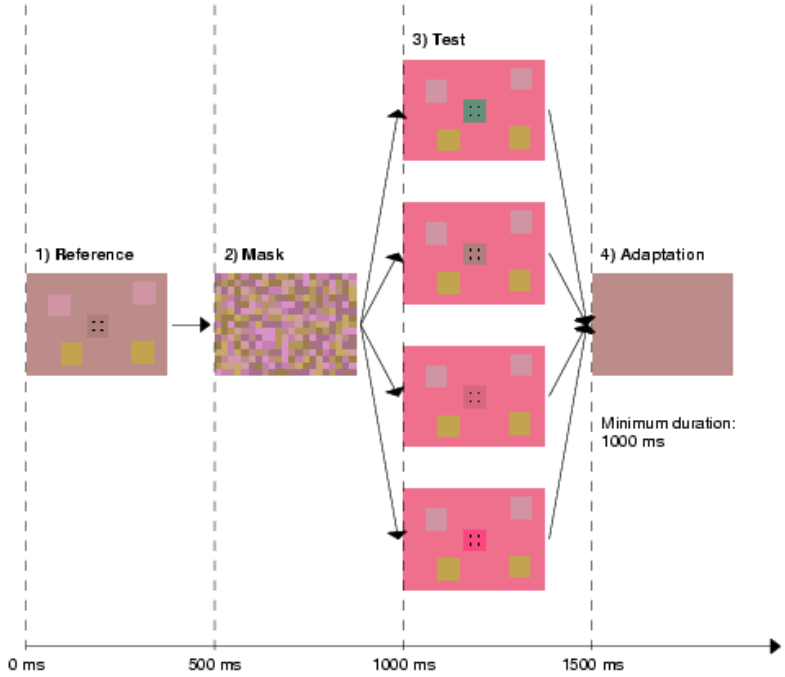


Figure 15: The paradigm used to measure simultaneous contrast induction. During the first 'reference' stage, the observer's task is to memorise the appearance of the 'reference' square against a neutral background. The observer views a mask composed of 0.25° squares that has the same space-averaged chromaticity and luminance as the background of the first stage. In the third stage, the central 'test' square may have any of six distinct chromaticities picked from an axis of constant S-cone excitation. The observer's task is to decide whether the test square appears 'redder' or 'greener' than the reference. Stage 4 is a short period of top-up adaptation, during which the observer signals his or her decision by pressing either of two gamepad buttons.

Measuring simultaneous contrast

The experimental procedure used by Wachtler et al. (2001) used a nulling method in combination with a 2 alternative forced choice task. Figure 15 shows the progression of a single trial using this method. In stage 1 of the trial, lasting 500 ms, a neutrally coloured reference square is presented against a neutrally coloured background[4]. The observer's task is to memorise the colour appearance of this reference square in this neutral setting, and compare it to the colour appearance of a test square in a new setting (stage 3 – multiple examples shown). During a short period of top-up adaptation (stage 4) the observer is asked to press a button to indicate whether the test square in stage 3 appeared 'reddish' or 'greenish' in comparison with the reference square in stage 1. A mask composed of small 0.25° squares was presented for 0.5 seconds between stages 1 and 3, its purpose being to ensure that colour appearance judgments were based on memory, rather than successive contrast.

Were the test square to retain the same chromaticity as the neutrally coloured reference square, most observers would judge that it became 'greenish' in appearance due to chromatic induction from the immediate background, that is red in stage 3. But by setting the colour of the test stimulus somewhere intermediate to the colours of the neutral reference and the strongly coloured inducing background, this shift in colour appearance may be neutralised. Using a method of constant stimuli, we determine the chromaticity of the test square that is judged 'reddish' 50% of the time. It should be emphasised that the colour shift of the background is constant within each block of trials.

[4] The choice of 'neutral' point is largely subjective. Possible choices include equal-energy white; the chromaticity of sunlight at noon etc. We took the chromaticity of our 'neutral' point from the colour used by Wachtler et al. It also serves as the whitepoint for later cone-conversion calculations.

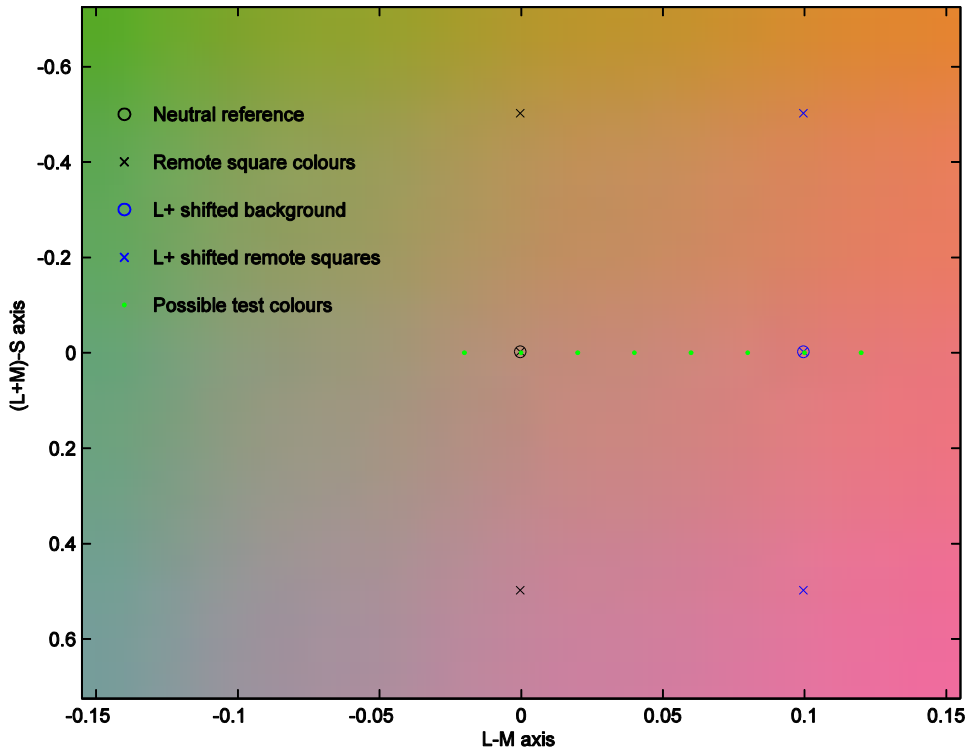


Figure 16: Stimulus chromaticities, plotted in an equiluminant cone-contrast space. The CIE 1931 xy coordinates of the reference chromaticity (chromaticity coordinates 0, 0) were 0.321 and 0.337 respectively. Chromaticity coordinates of the reference background and remote-surfaces are plotted in black; L+ shifted coordinates of the same surfaces are marked in blue. Some possible 'test' chromaticities of the central square are marked as a line of green points, though in practice colours for this stimulus element could be chosen from any point lying on this axis.

Figure 16 approximates the modified MacLeod-Boynton colourspace that we used in the experiment, with the stimulus colours that we used plotted against it. The line of green dots marks the axis in colourspace from which test chromaticities could be taken, for the method of constant-stimuli that we used.

The strength of simultaneous contrast induction is expressed as the proportion of the background chromatic shift that must be applied to the central test square if it is to retain its neutral appearance in its new setting. A high score indicates that the colour of the test-square is strongly dependent on that of the background. A low score indicates that its colour appearance is primarily determined by its chromaticity, and that simultaneous induction is weak in a given setting.

This task is very different from tasks used by most previous experimenters, such as Jameson and Hurvich (1960) or Chichilnisky & Wandell (1995) who used consecutive or simultaneous matching methods, where observers memorised the appearance of a coloured patch against a strongly coloured background, then either identified a similarly coloured chip from a selection of Munsell-patches, or adjusted a computer-generated stimulus to match the colour of the original. These experiments were typically carried out under free-viewing conditions, where observers were not time-constrained in making their choices. It also differed from the dynamic nulling technique employed by Quasim Zaidi et al., (Zaidi, Yoshimi et al. 1991; Zaidi, Yoshimi et al. 1992) where the colour of the background oscillated along an opponent axis, and observers adjusted a constant relating the chromatic shift of a test center to the instantaneous chromatic shift of the background, relative to a neutral reference.

Amongst its strengths, the paradigm of Wachtler et al. (2001) may count the fact that individual trials are limited in duration, reducing the effects of long-term adaptation and maximising the influence of short-term simultaneous contrast effects over the results. Any 1-dimensional nulling technique is subject to the criticism that the observer may never be able to find an exact match for the neutral reference colour. However, our experiment does have one advantage over other nulling techniques in that the observer only has to make a single match that is valid for each background change that is studied, whereas in continuous nulling experiments such as those carried out by Zaidi and others (Zaidi, Yoshimi et al. 1991; Zaidi, Yoshimi et al. 1992; Zaidi and Zipser 1993), the observer is effectively matching an infinity of background colours, with a corresponding infinite set of target chromaticities. Furthermore, we are accustomed to making individual judgments about colour-matches in our everyday lives, whereas setting a match in a nulling experiment is a more artificial task.

There are, of course, some unknowns in the experiment. Rinner and Gegenfurtner have described the time course of adaptation to DC steps, showing that it is mediated by at least three mechanisms, with a different time constant in each case (Rinner and Gegenfurtner 2000). However, adaptation also occurs to reduce our sensitivity to either repeated consecutive or simultaneous chromatic contrasts, and the time-course of this process (or in all likelihood, these processes) remain relatively unstudied, as far as I am aware.

Stimuli

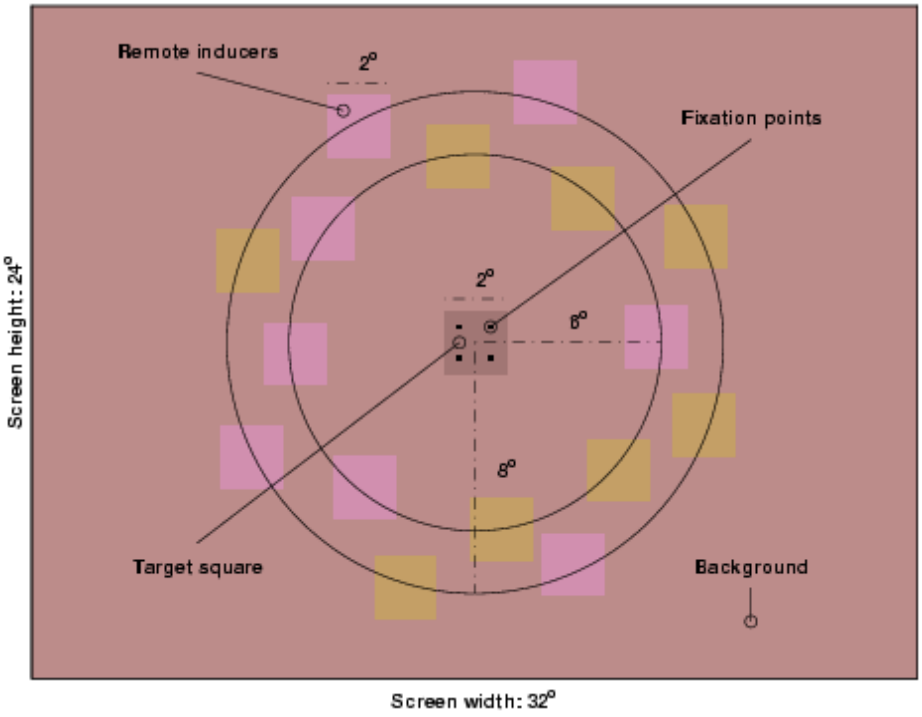


Figure 17: The main features of a prototypic stimulus as used in the condition with only two distinct remote surface chromaticities, drawn to scale but with uncalibrated colour reproduction. Stimulus brightnesses and chromaticities varied between conditions and between the different stages of each trial: for details see the text.

The full-screen background subtends 32° x 24° of visual angle, and a 2° central target whose colour must be memorised or judged, is positioned at the center of the display. Four small (0.25°) black points, visible during the first three stages in each trial, help the observer maintain fixation. The remote-inducers are loosely arranged into two rings surrounding the target square; each remote surface is displaced by up to 0.5° in a random direction to ensure that their positions differ in each trial and to make their arrangement appear less 'regular'.

In experiment 3 the monitor was divided centrally, each half of the screen being visible to one eye only. Fewer remote-inducers were used as the display area available was halved, but the overall stimulus configuration was similar. Stimulus chromaticities differ between experiments, and are detailed within the text.

Figure 17 shows the prototypic stimulus that was used and modified throughout our experiments. It consists of a central square (size 2°; CIE co-ordinates x = 0.321, y = 0.337) surrounded by two concentric rings of remote squares (size 2°) 2°, 6° or 8° distant, all against a 32° by 24° uniform, neutral background (CIE coordinates Y = 18.7 Cd/M^2, x = 0.321, y = 0.337). In different stages of each trial, the colours of the background and target square were modulated as previously described. The colours of the remote surfaces, differed from the colours of the neutral and L+ shifted backgrounds only in S-cone contrast - figure 16 shows the chromaticities of the surround for the experimental setting when only two distinct chromaticities were used.

I have described above the method that we used to investigate the strength of simultaneous contrast. In each part of this experiment, we determined the strength of simultaneous contrast for four different conditions:

1) No background change; no remote change (*"no change"*)

2) No background change; remote L+ shift (*"remote change"*)

3) Background L+ shift; no remote change (*"local change"*)

4) Background L+ shift; remote L+ shift (*"global change"*)

Figure 18 illustrates these four conditions graphically, and figure 19 shows how the results were analysed and compared across these four conditions, by fitting a Weibull function to the percentage of test squares that were judged reddish, against their L-cone contrasts relative to the neutral reference stimulus.

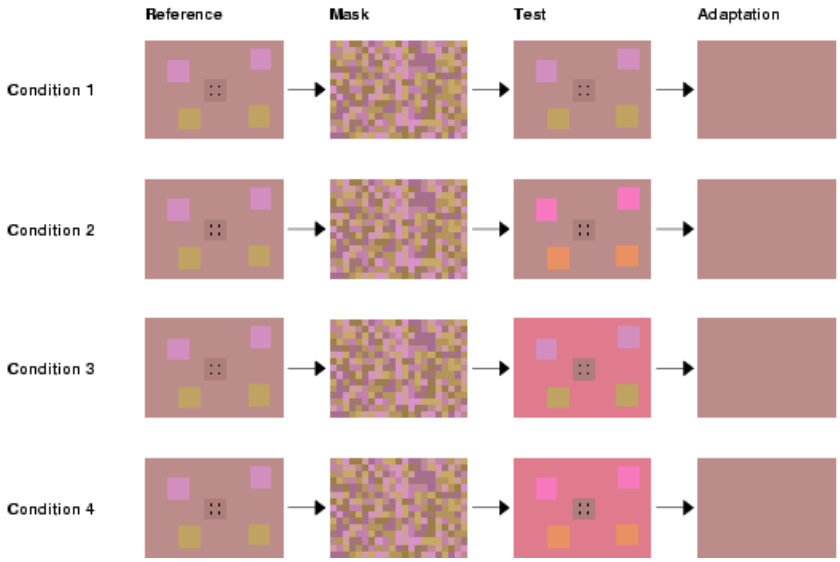


Figure 18: Four conditions were run for each experiment. Each condition was identical, save for the third 'test' stage, in which the chromaticities of the background or the remote inducers could be shifted in the direction of increasing L-cone contrast. Four conditions were therefore possible:

1. *No shift for either background or remote squares.*
2. *Shift in remote squares only (" remote change").*
3. *Shift in background only ("local change").*
4. *Shift of both background and remote squares ("global change").*

This figure is not to scale and does not show all of the remote surfaces.

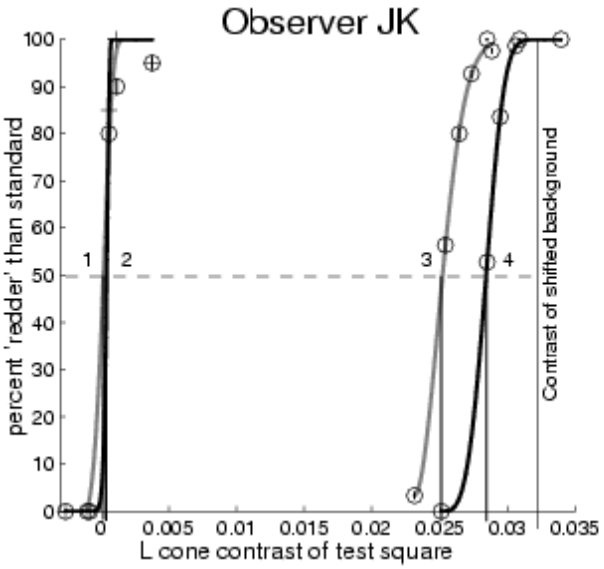


Figure 19: The percentage of responses in which the test square is judged 'redder' than the reference square are plotted against the L-cone contrasts (modified MacLeod-Boynton coordinates) of the test, relative to the neutral background, for observer J. K. Smooth curves are best-fitting Weibull functions. Numbers 1-4 label the curves corresponding to the 4 conditions described earlier. Each data point is the average of at least 45 trials.

As can be seen on figure 19, the leftmost two curves closely overlap, but a pair of curves further to the right of the stimulus are more widely spaced. The first two curves are from the conditions in which the background colour did not change, this indicates that the remote backgrounds did not influence the perceived colour of the target square in the absence of a background colour shift. However, in the presence of a background colour shift, curves 3 and 4 do not overlap indicating that the colour perceived as identical to the neutral reference differs in each case. This is the key effect that Wachtler et al. (2001) studied, and that we described as the 'contrast-enhancement' effect central to our paper (Wolf and Hurlbert 2002) – i.e. contrast induction is stronger for condition 4 compared with condition 3. However, we never investigated the extent of contrast-induction for these stimuli in the absence of the remote inducers, so it may be equally valid to view it as a 'contrast-diminution' effect, where contrast induction is weakened in condition 3 relative to condition 4. Henceforth in this chapter, I refer to it as the remote-contrast-modulation-effect.

Stimulus generation and presentation

Stimuli for this experiment were generated using a Silicon-Graphics Indy workstation using an indexed 256 colour mode. They were displayed on a 17" Silicon Graphics monitor, and viewed from a distance of 57 cm through a viewing box to which a removable septum and mirrors could be added, to form a Wheatestone stereoscope. In experiment 1.3 (binocular viewing) the Wheatestone apparatus added an additional 10 cm to this distance, causing the stimuli to appear around 15% smaller.

Chichilnisky et al. lined the side of their viewing box with mirrors in order to ensure stable adaptation (Chichilnisky and Wandell 1995). Without such mirrors only the central portion of the observer's visual field is filled by the computer screen; eye movements may cause some areas of the retina to be directed alternately to the computer screen then to the darkness surrounding it. Mirrors were not suitable for our experiment, because they would have produced multiple virtual images of the stimuli, akin to a kaleidoscope. We chose instead to line our viewing box with white paper, whose reflectance we determined to be around 95% throughout the visible spectrum. The paper's chromaticity was uniform and similar to the mean chromaticity of the computer screen. However, its brightness was lower than that of the computer screen and was not uniform. The net result was somewhat akin to looking into a Ganzfeldt dome, though less immersive as the view of the paper did not completely fill the peripheries of the visual field.

Experimental organisation

Three observers were recruited by advertising around the University for observers, and my supervisor and I also carried out the experiments. Before starting to collect experimental data, new observers learned to perform the task by practicing the easiest condition – neutral, uniform-coloured backgrounds. As they became adept at this, and started to have a clear idea of what they were doing, they practiced with neutral articulated backgrounds and finally moved on to uniform and articulated colour-shifted backgrounds, where the colour-judgments are more difficult. The judgment of when they were ready to start collecting data was made arbitrarily based on the consistency of their data.

For each condition four separate blocks of trials had to be run: two for the neutral trials (with and without remote inducer shift), and two for the shifted-background trials (with and without remote inducer shifts). These could not be mixed at random, as this would cause changes in the observer's adaptational state over the course of the experiment. During any single block of trials, the only parameter to change from trial to trial was the mean chromaticity of the target square.

We found that over time, the 'neutral point' chosen by an observer moved along the inducing axis. Typically this caused the strength of the contrast induction/diminution effect to decrease in magnitude over the course of the experiments. Given the smallness of the effect we were measuring, it was important to avoid any influence this may have had over the results. For this reason, conditions that are compared within a single graph were run in close temporal succession. Blocks measuring different conditions were mixed within a single observing session, and we attempted to collect full datasets as quickly as possible. Conditions that were identical but used in more than one graph were not necessarily repeated, provided they had been measured sufficiently recently. This effect may therefore influence variance, but should not bias the results.

The number of trials for each data point is quite variable. Again, I collected as many points as possible, for the full range of conditions. For the other observers we reduced the number of data points in order to reduce their observing time as far as possible, whilst still obtaining sufficiently small error bars for the null point.

Observing sessions lasted between 60 and 90 minutes, with individual blocks generally lasting around 10 minutes. Observers had a short break between each block of tests, and at the half-way point they were encouraged to take a 5-10 minute break and leave the tunnel. The number of trials within a single block could be set arbitrarily, and it was often necessary to run short blocks in order to verify that the target square test chromaticities were correctly sited around the observer's threshold. Were these parameters found to be correct, the data would be included in the results, even if the block of trials was relatively short. Conversely, longer blocks were sometimes run at the end of a session, when there was insufficient time to run two normal-length blocks. The number of trials per point[5] per block could vary, depending on the number of distinct conditions being measured and the density with which the psychometric function was being sampled. If observing sessions ended prematurely – for example in the case of a fire alarm – all data collected up to that point could still be utilised in the analysis, so even within blocks the number of trials per point was not always equal. For the neutral background conditions, where observer performance was superior, a minimum of 20 trials per datapoint were collected. For conditions with shifted backgrounds, a minimum of 30 trials per point were taken. In general, however, around 45 trials per point were presented. For some conditions, observer KW measured up to 110 trials per point.

Further details on the age, sex and colour vision of the observers are given in Appendix A.

Statistical analysis

Our threshold fitting program, written by Bruce Cumming and described in the Methods chapter, chapter 3, calculates the standard error of the mean on the threshold. Where an error bar on a difference between thresholds is reported, as in the figures above and below, the error is calculated as the square root of the sum-of-squared-errors, on the s.e.m.s of the respective thresholds. All displayed error bars are therefore either standard error of the mean, or the combined standard error of the mean.

[5] By 'trials per point' I refer to the datapoints used to sample the psychometric curve, such as that illustrated by Figure 19.

As this experiment employed the 'method of constant stimuli', the statistical analysis was complex and it was not possible to employ ordinary ANOVA to analyse the data. An additional problem, not shared by experiments 2 and 3, is that the data bars shown in this chapter themselves represent comparisons. In performing significance tests to determine whether one bar differs from another, we therefore need to find a method capable of determining the significance of comparisons of comparisons. This is something that neither Wichmann and Hill's 'pfit' method, nor the chi-squared method employed by Wachtler et al. (2001), would be readily able to do.

In interpreting these experiments, the two options that are available to us are to interpret the significance of differences based on the error-bars, which in all cases show the standard-error of the mean. The second alternative, is to examine the replicability of the results between observers. If most of the observers show the same pattern of data, with sufficiently small error bars relative to what the data show, then it should be valid to accept any findings despite the lack of formal statistical analysis..

Ideally one should perhaps also have analysed the blocks of data to determine whether the observers are subject to any learning effects. Given the difficulty in analyzing these datasets with standard statistical tools, I have not attempted this. However, during the course of the experiments I plotted the data to determine whether there were any consistent threshold shifts between observing sessions. Only small shifts were noted, well within the bounds of the error bars. The design of the experiments where any conditions that are directly compared were measured during the same observing sessions, should also have served to negate any bias due to threshold drift over time.

Results

Experiment 1.1

Here we test the hypothesis that increasing the number of distinct surfaces that change consistently with the global illumination change, affects the magnitude of the contrast-modulating effect. We compared the effects of two different remote colour distributions, using luminances of 18.7 Cd/M^2 for the background and a relative luminance decrement of -20% for the test patch and an increment of +12% for the remote fields. In the 2 colour distribution, the remote square S-cone contrasts were ±0.5 (8 squares of each), and in the 8 colour distribution, ±0.5, ±0.375, ±0.25 and ±0.125 (2 squares each), relative to the relevant background.

Our prediction is that when more distinct chromaticities undergo an identical illuminant shift, our assumption that the background change is an illuminant shift rather than a surface shift will be reinforced and the strength of simultaneous contrast induction should therefore be increased. This should result in positive values for the contrast-modulation effect, with the 8-colour condition bars being more strongly positive than the 2-distinct colour condition.

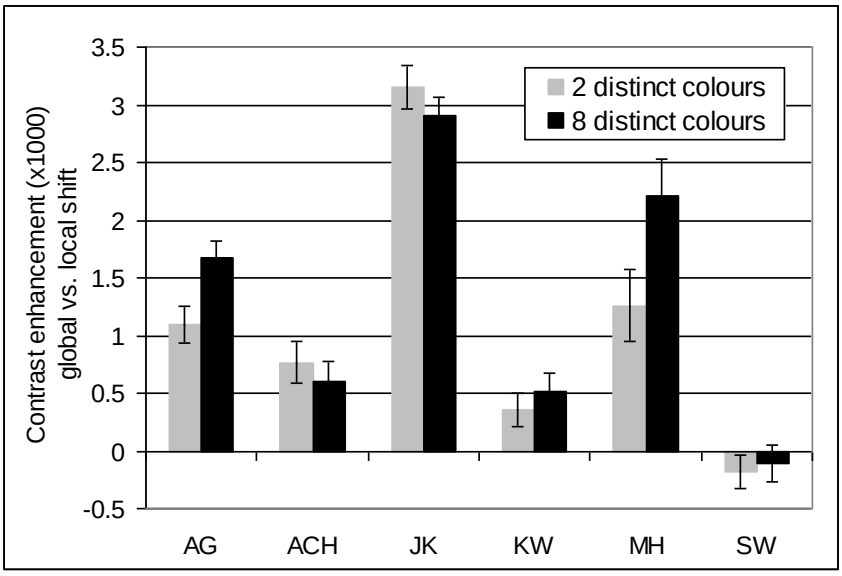


Figure 20: The magnitude of the contrast modulation effect for each observer, plotted as in Figure 19. Filled bars show the difference in L-cone contrast (x 1000) of the perceptually neutral square (50% point on the fitted Weibull function) for the global change versus local change conditions ("contrast modulation"). Gray bars: 2 distinct colour condition. Black bars: 8 distinct colour conditions. Error bars are combined standard errors of the mean.

Figure 20 shows the results from this experiment. Contrary to our expectations, the 2-colour and 8-colour conditions were equivalent for all but two (out of six) observers, for whom the 8-colour condition increased the contrast modulation caused by a global change. This result suggests that the mechanism responsible for contrast modulation on the L-cone-axis may be insensitive to variation solely along the S-cone-axis. In other experiments (results not shown) we varied the remote chromaticities along the L-cone-axis, and shifted only the S-cone-contrast of the background. There we found no global contrast modulation effect in all four observers tested (contrary to Wachtler et al. 2001), again suggesting that long-range chromatic interactions differ qualitatively across the cardinal axes (Cornelissen and Brenner 1991; Barnes, Wei et al. 1999).

Experiment 1.2

Here we test the hypothesis that increasing the contrast between the surfaces, thereby indicating that they form parts of distinct illumination frameworks, reduces the magnitude of the remote contrast modulating effect. Two conditions are compared – a high contrast condition and a low-contrast condition. The 'high contrast' condition is identical to that used in experiment 1.1 – the test patch is 20% dimmer than the background and the remote fields are 12% brighter. In the low-contrast condition, the test patch is 7.5% dimmer and the remote fields are 3.5% brighter. Only two distinct colours are used for the remote fields, created by modulating their S-cone contrast by +/- 0.5 relative to the background chromaticity.

Small changes in perceptual grouping may appreciably affect lightness perception (Adelson 1993) and colour appearance (Schirillo and Shevell 2000). Here, ungrouping the stimulus elements may reduce the importance that global coherence plays in determining whether a colour change is due to a change in illumination or surface material. Hence we would predict that the contrast modulation effect would be weaker for the high luminance contrast condition than for the low luminance contrast condition.

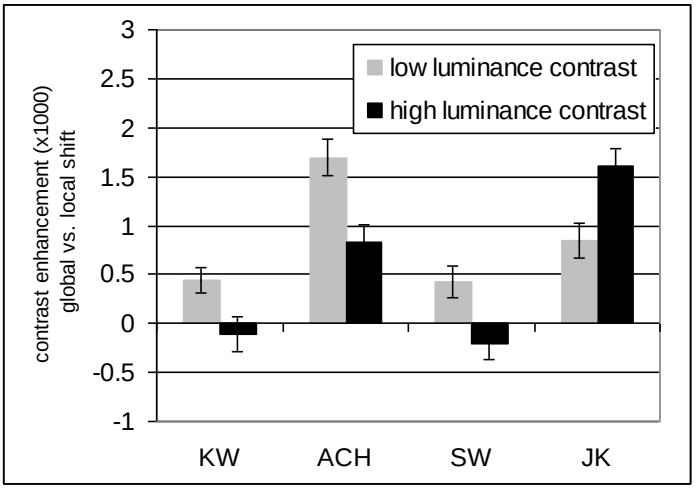


Figure 21: Filled bars show the difference in L-cone contrasts (x 1000) of the perceptually neutral square (50% point on the fitted Weibull function) for the global change versus local change conditions (" contrast modulation"). Gray bars: low luminance contrast condition. Black bars: high luminance contrast condition. Error bars are the standard deviations of the difference.

Figure 21 shows the results of this experiment. For all but one observer (JK), increasing the luminance contrast between the different stimulus elements decreases the magnitude of the remote contrast modulating effect. This is consistent with our working hypothesis.

Experiment 1.3

In this experiment, we aimed to determine whether the remote fields exerted their influence at a monocular or binocular site in the visual system. We therefore compare conditions where we present the remote fields to the same eye, or to a different eye, from the test field. Here, we make no predictions as to the likely outcome – this is not a hypothesis-testing experiment, but a 'critical test'.

For the haploscopic experiments, we used a smaller stimulus (16° by 24° in each eye) with only four remote inducers (of 2 colours, ±0.5 S-cone contrast; low luminance contrast condition) 4° distant from the center. This change was made simply for logistical reasons as only half the screen was available to each eye. The fixation dots were presented to both eyes to stabilise convergence. With the exception of one observer, who was excluded following a test-run of the experiment, all observers reported fusion of the combined stimulus to be straightforward and stable. In each session, the central square was presented to one eye only, but the uniform background (either neutral or L-shifted) was presented to both. The results of right-eye and left-eye sessions were then pooled for analysis. Three separate conditions were interleaved in each session:

1. remote squares presented only to same eye as central square (`same eye' condition
2. remote squares presented only to the opposite eye (`different eye' condition)
3. remote squares absent on second stimulus presentation

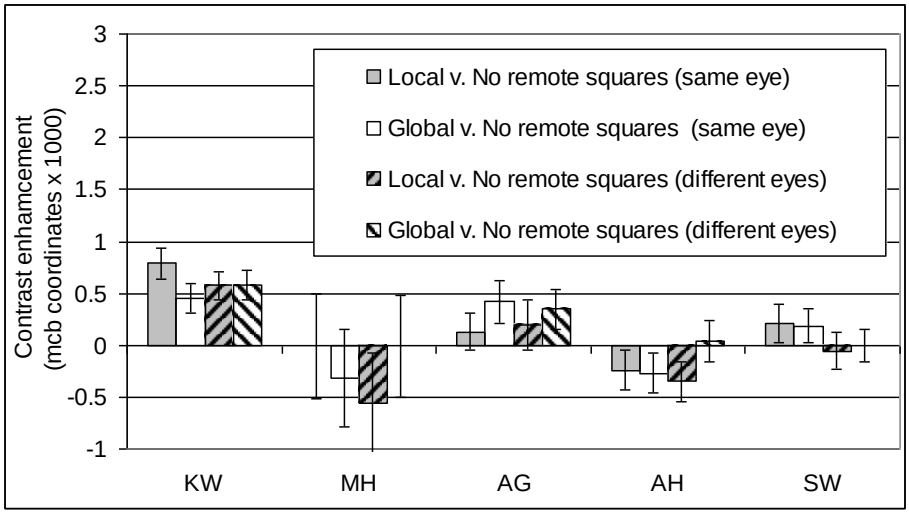


Figure 22: Contrast modulation effects for Experiment 3 measured as the changes in the perceptually neutral L-cone contrast (x 1000) caused by local or global shifts plotted relative to the reference threshold for the no-remote squares background shift. Solid bars: same eye. Hatched bars: different eye.

Figure 22 shows that the contrast modulation effect disappears or is greatly reduced for all observers under haploscopic viewing. Whether the surrounding squares are presented to the same or separate eyes, the differences between the neutral thresholds for the global and local shift conditions are minimal. Binocular interactions are unlikely to alter the chromatic appearance of the remote squares themselves, as such effects are small when there are no aligned borders in the two eyes (Whittle and Challands 1969; Shepherd 1997). Hence, it seems likely that the uniform background presented to the opposite eye overrides induction by the remote squares. As either explanation requires interaction at a binocular level, the remote contrast modulation effect is unlikely to be monocular.

Discussion

What significance might the phenomenon from this experiment have in a natural setting? A common objection to these experiments is that the task is not very *'natural'* – it is difficult to think of everyday parallels to the task employed here, in marked contrast to the consecutive comparison tasks previously outlined which are similar to those we perform when we judge food colours, pick out cosmetics or choose matching fabrics. Nascimento and Foster have suggested that rapid illuminant changes are found in nature (Nascimento and Foster 2001), for example when a passing cloud blocks light from the sun – though such changes are unlikely to be instantaneous as the chromatic shifts in our stimuli were. Additionally, the eye may also make very quick saccades between areas of a scene that are under different illuminants, though this situation also differs from the scenario in our experiments, where stimulus colours are changed without regard to eye movements, or even in the absence of saccades.

The stimuli used in this experiment were probably not designed to replicate or even closely resemble any particular situation found in nature, and in my view they are best viewed as a highly artificial means of investigating contrast gain-control mechanisms. Yet there is some evidence that the tasks we are asked to do for psychophysics experiments can qualitatively influence the results obtained. Arend and Reeves found that people made different matches if they were asked to pick either *'the paper that the original object would have been made from'* or the *'stimulus with the same appearance as the original'* in a simultaneous matching task (Arend, Reeves et al. 1991). In a separate experiment, Bramwell and Hurlbert found that colour matches set by a method of adjustment differed from matches determined by a method of constant-stimuli (Bramwell and Hurlbert 1996). These findings provide strong evidence that our performance at psychophysical tasks can be subject to higher influences even when the tasks and stimuli appear simple in nature.

Wachtler et al. (2001) were aware of this possibility and asked their observers whether they consciously interpreted the stimuli to be representations of real objects – such as shapes cut from paper. It is reassuring that all the naïve participants in his experiment seemed to interpret the experiment in the same way and reported that they did not consider the stimuli as real objects. This may be taken as evidence that the responses were made purely on the basis of hue judgments, however it is also possible that the increased colour-constancy index reported by Hurlbert and Bramwell was due to differences in experimental strategy used by the observers rather than to a duality of perception.

Rather than asking 'what natural situation do these stimuli replicate' one may ask what theoretical utility the contrast gain control mechanism that is the focus of these experiments, may have. The two primary roles that have been proposed for chromatic contrast are in colour constancy, and as a gain-control mechanism for maximising the salience of chromatic variation. Either role may call for some form of contrast-gain-control to be exercised, though it is possible that some alternative explanation exists. For example, it may be that contrast induction serves no useful role in the natural world, but is simply an epiphenomenon of the way in which the visual system is constructed.

In view of this, the results obtained by this method must not be over-interpreted. They can provide evidence for the existence of neural mechanisms able to perform particular functions, but for the present it is unclear what role these phenomena have outside of the laboratory.

Conclusions

Our firmest finding was to replicate the basic effect described by Wachtler et al. in their 2001 paper. Whilst it was small and its magnitude varied strongly between observers, our data provide independent confirmation of its existence in a total of five subjects.

Could the effect stem from some artifact, common to the computer systems used both by ourselves and Wachtler's group? Mechanisms by which such effects could occur are certainly feasible. As shown in the methods chapter, spatial dependencies with this magnitude can easily occur. Yet, the bugbear of the experiment – the large amount of inter-individual variation – also makes this scenario unlikely, as if this were true the artifact should manifest itself in the same way for all the observers. Might the artifact might only be apparent whilst the monitor was warming up or if the observation tunnel was at a particular room temperature? Despite varying the subject timetable, there was no perceptible change in the pattern of the results therefore I believe the contrast-gain-control effect studies by ourselves and Wachtler et al. is probably real.

Another possibility is that the observers somehow differed in the method by which they performed the experimental task. We had no method of monitoring subjects' eye movements so it could be that some observers made furtive glances to the periphery during the test stage of the experiment. If so, foveal adaptation to the remote inducers could account for the measured reduction in induced contrast, though it would still be necessary to explain why this is only observed in conditions 3 and 4 when the local background also changes. We did no experiments to control for eye movements and had no method of monitoring them, but Wachtler has shown that his effect persists even when the stimulus duration is reduced until it is too brief for saccades to occur (personal communication). Yet another tentative explanation for inter-individual differences is that the magnitude of the effect depended on the amount of attention observers directed towards the properties of the surround.

On reflection, the two chief characteristics of the effect are that it is small and highly variable between observers. Does it then have any practical, rather than merely theoretical, significance? It is certainly possible that Wachtler's effect is a weak demonstration of a phenomenon that is important in determining the colours that we perceive. But despite much effort neither he nor I ever found a more effective way of eliciting it. It was difficult, expensive and time consuming to study as I did not show the effect I could not pilot or debug new versions of the experiment. We also had to collect a great deal of data in order to ensure that the error bars were small relative to the contrast modulation effect. Ultimately I decided to move on in search of more fertile pastures.

Chapter 6, Experiment 2 – Texture

Background

Texture segmentation and contrast induction

Gestalt psychologists have identified a number of segmentation cues that can be used to determine whether elements are part of the 'figure' or the 'ground'. These include texture or feature-based cues, perceptual grouping, motion and depth. Other cues, such as brightness can also serve to divide scenes into closely related and independent parts.

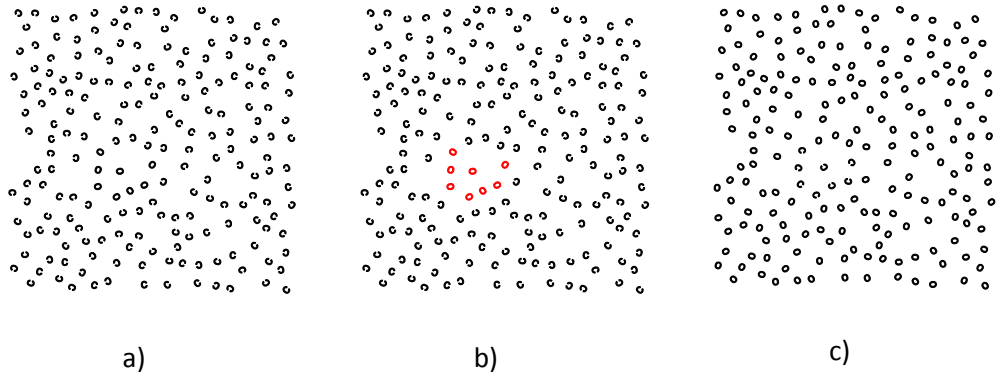


Figure 23: Asymmetric figure-ground segmentation

a) shows a small number of 'O's hidden amongst a sea of 'C's. Their position is not immediately apparent.

b) shows the same pattern as (a), but highlights the 'O's in red.

c) shows a small number of 'C's hidden amongst a sea of 'O's. In contrast to the 'O's in (a), their position is readily apparent, illustrating the phenomenon of 'asymmetric popout'.

Texture cues can provide strong 'popout' cues, segmenting figure from ground. Figure 23 shows how elements within a texture can cause certain regions of it to appear immediately distinct from the background – whilst apparently similar changes may not have the same effect.

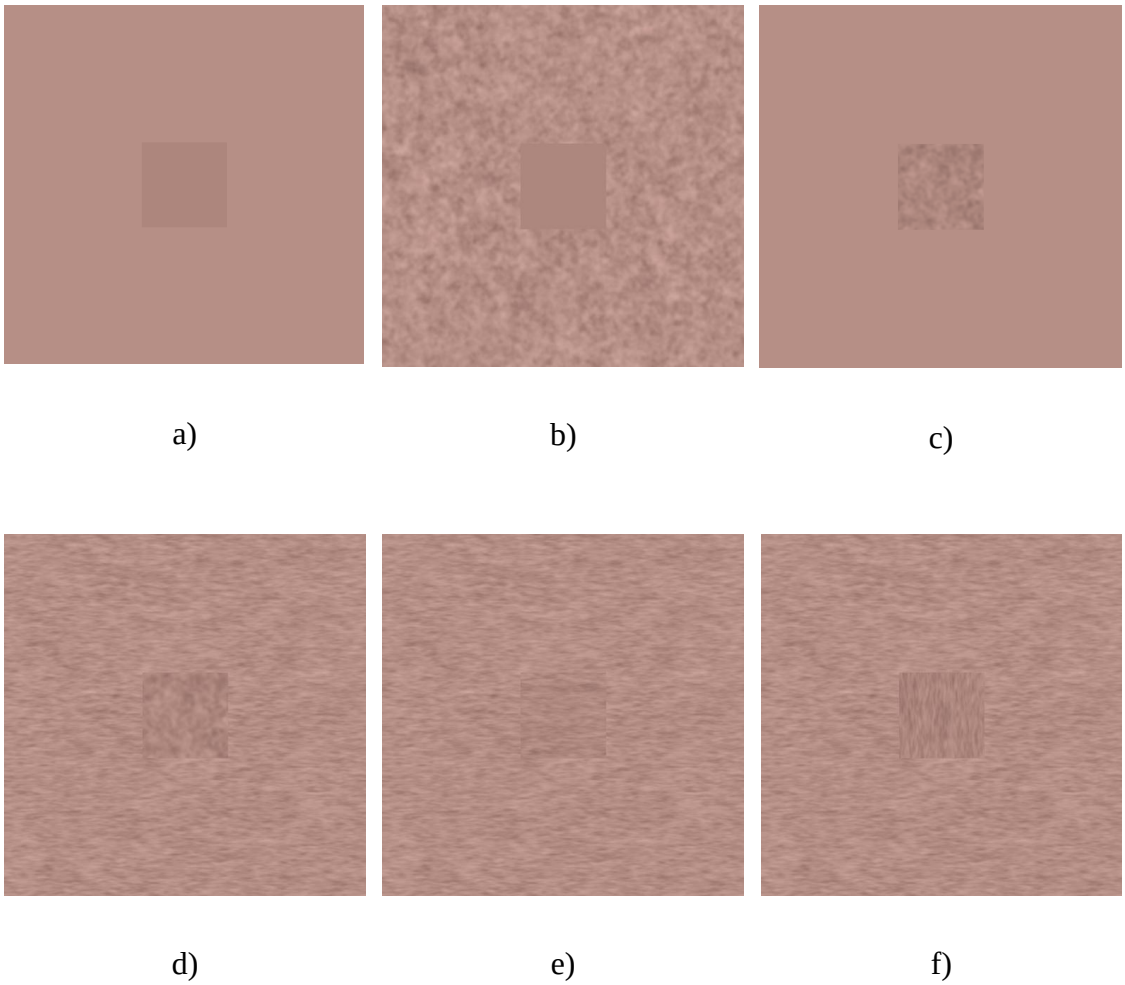


Figure 24: Figures and grounds – or targets and surrounds – with various component textures

The mean brightnesses of the six preceding textures are identical in each case. These textures were prepared by performing a 2d Fourier-transform on random noise, and rescaling the frequency components prior to performing the inverse transform.

a) Uniform target and surround

b) Textured surround; uniform target

c) Uniform surround; textured target

d) Anisotropic surround texture; isotropic target texture

e) Anisotropic surround and target – identical texture orientation

f) Anisotropic surround and target textures – different orientation

Different combinations of texture in the targets and surrounds cause them to appear distinct to differing degrees.

Figure 24 provides further examples, showing how texture can also drive segmentation. A range of figure-ground arrangements are shown, with different arrangements of luminance texture in the background and surround. In each example, the mean lightness of the figure and ground is identical. But in some arrangements the figure-ground relationship is more striking than in others. The figure in figure 24e is probably less distinct than the figure in figure 24a, but if the anisotropic texture in the figure is rotated 90^{o} relative to that of the background as in figure 24f, it re-emerges strongly from the background. The textures in these figures were generated by filtering noise, meaning that their spectral properties are well known and controlled, but that they are somewhat unadventurous compared to natural textures such as stone, bark or skin.

Aims

The aims of the experiments discussed in this chapter, are to examine how visual 'texture' affects simultaneous contrast induction. My rationales for embarking on this set of experiments were twofold. Firstly, I wished to investigate whether simultaneous contrast induction is affected by cues to segmentation. Secondly, I wished to determine whether it was possible to use the 2-AFC method to investigate our perception of polychromatic objects – that is, objects whose surfaces are composed of more than a single component colour.

My initial results demonstrated that texture affects simultaneous contrast very strongly. I therefore split the experiments described in this chapter into two categories: my initial set of experiments, designed simply to determine whether texture influences simultaneous contrast induction, and follow-up experiments designed to examine why and how simultaneous contrast is affected by adding texture to my stimuli.

The following headings give a brief overview of the aims of each set of experiments. Their aims are expounded in more detail in their relevant subsections.

Does texture influence simultaneous contrast induction?

2.1.1 Does texture in the inducing background influence simultaneous contrast induction?

2.1.2 Does contrast induction act differently on textured targets?

Why and how does texture influence simultaneous contrast induction?

2.2.1 Does texture affect contrast induction due to high-level segmentation, or some other factor?

2.2.2 Do textured surfaces affect contrast due to their global properties, or is the effect of texture due to its action at boundaries between objects?

2.2.3 Is the influence of texture in a surface in proportion or independent of its contrast?

2.2.4 Is induction tuned to the cardinal chromatic axes?

Methods

Measuring simultaneous contrast

All the experiments discussed in this chapter were based on a 2-alternative forced-choice paradigm that evolved from the method used in experiment 1.

Trial structure

Figure 25 illustrates the structure of an individual trial in these experiments. A short (500 millisecond) period of top-up adaptation is shown, during which the observer signals his response from stage 3 by means of a gamepad. On each trial the following sequence was presented:

1. reference stimulus (500 ms)
2. mask (500 ms)*
3. test stimulus (500 ms)
4. neutral background / top-up adaptation (minimum 500ms).

* The mask is not shown in all versions of the experiment

The test stimulus was identical to the first in its spatial configuration; only the colours differed. For the basic experimental setup, the background colour or colours were uniformly translated along the (L-M) axis in the isoluminant plane relative to their prototypic colours, always with a constant total cone contrast shift of 0.1. The total cone contrast of the central square was varied between trials, its value taken at random from a set of constant increments along the isoluminant (L-M) axis relative to the prototypic square colour. In some experiments we also investigated chromatic shifts along other chromatic axes; further details are given on an individual basis.

In L-M contrast shifts, the observer's task was to indicate by a button-press whether the test square appeared 'redder' or 'greener' than the prototypic square. Responses were taken during the top-up adaptation period (minimum 500ms) between trials. The observer was exposed to the neutral illuminant for at least 2/3 of the duration of each block of experiments. However, if he was slow to respond to the stimuli, this proportion could increase because the increased relative duration of the top-up phase.

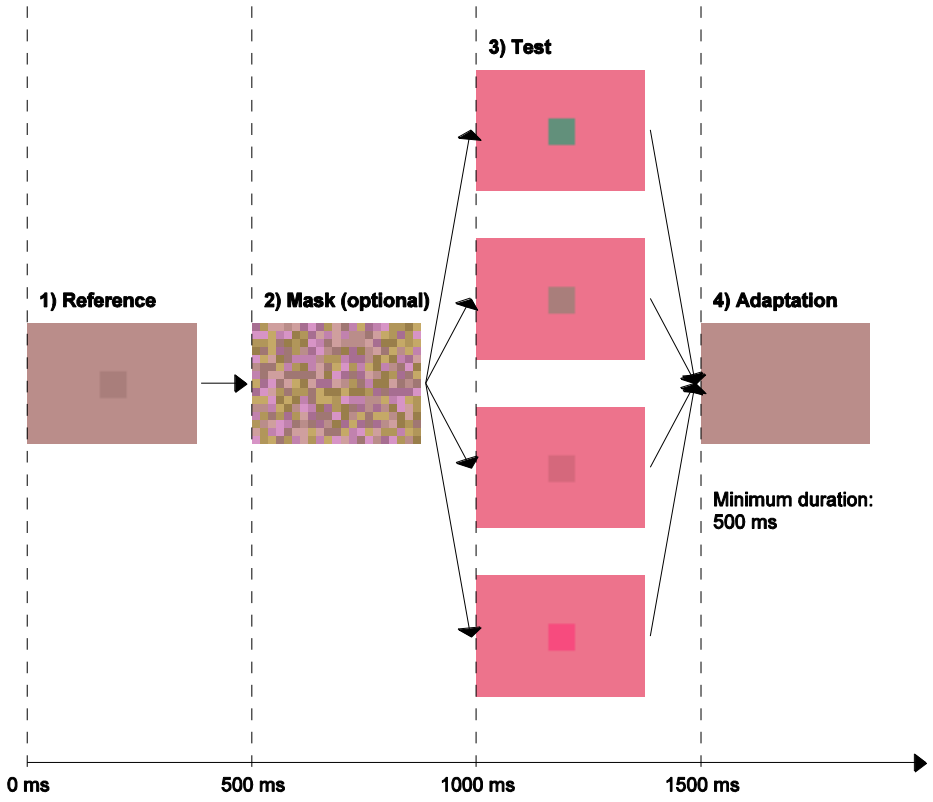


Figure 25: The basic experimental paradigm used for all the experiments in this chapter

1. *In the first stage, a reference target is presented, set against a neutral background. The observer fixates on this square, and memorises its colour for future reference.*
2. *In some experiments, a mask was shown. However its use does not seem to affect the results, so in order to increase the efficiency of the experiments I removed this stage from later experiments.*
3. *A test target is presented against an L+ shifted background that is equiluminant with regard to the first background. The background shift is constant, so the same reference and L+ shifted backgrounds are used throughout the experiment. The colour of the target square may range between red and green, taken from a chromatic axis passing through both the chromaticity co-ordinates of both the reference and L+ shifted backgrounds. The observer's task is to judge whether the target square appears redder or greener in this condition, with respect to the reference target.*

In this figure, only uniformly coloured surfaces are shown. Details of the texture, chromaticities and luminances are given individually for each experiment.

Controlling adaptation

To enforce pre-adaptation, no responses were collected for the first two minutes of each experimental session; target colours were chosen at random from those available for other parts of the experiment.

Each psychometric function was fit by a logistic function, determined by the program 'psychofit', which estimated parameters including:

1) The point of subjective equality (point where the curve crosses the 50% mark)

2) The slope of the logistic function

3) Confidence intervals for both of the above.

We quantified the chromatic induction for each condition as the total cone contrast of the central test square (relative to the prototypic square's neutral colour) which, when viewed against the L+-shifted background, appeared neither redder nor greener than the prototypic square. This value was calculated as the 50% point of the best-fitting logistic function for the psychometric data relating percentage 'redder' responses to total cone contrast of the test square (see figure 26).

Data analysis

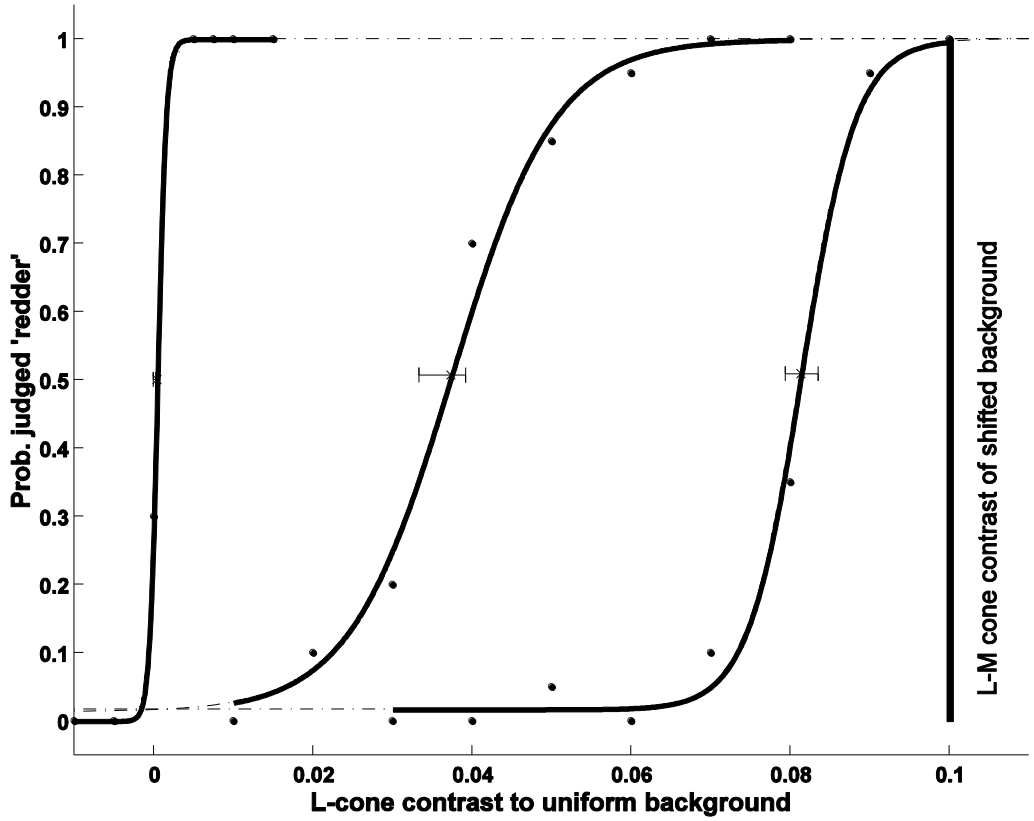


Figure 26: Psychometric functions for three conditions, for a typical observer. Filled symbols show the proportion of trials for which the observer judged the central test square to be 'redder' than the prototypic square, plotted against the test square's total cone contrast relative to the prototypic square's chromaticity. Smooth curves show best-fitting logistic functions, and horizontal error bars show 95% confidence intervals for the perceptually neutral contrast.

Leftmost curve: Control condition: uniform background; no shift in chromaticity between prototypic and test backgrounds.

Middle curve: Articulated background (9 squares-per-degree); L+ shifted test background. An asterisk marks the total cone contrast (0.1) of the test background relative to the prototypic background.

Rightmost curve: Uniform background; test background L+ shifted in chromaticity relative to the prototypic background, as for the middle curve. Asterisk marks the total cone contrast (0.1) of the test background relative to the prototypic background

Each data point is the average of at least 20 trials.

Inferential statistics

Statistical data analysis for data produced by the 'method of constant stimuli' is difficult to perform. Appendix D provides the results of a literature survey, examining whether other researchers analysed their 'method of constants' data using inferential statistics, and if so, what methods they used. Whilst two papers were found that employed ANOVA, they did not provide further details on their method. Appendix E includes the results of some Monte-Carlo simulations I performed, showing how slight changes in how ANOVA is performed can produce very large changes in the resulting p-values. Alternatives to ANOVA such as the Monte-Carlo method 'pfcmp', provided as part of the 'psignifit' package, produce more reliable results but are limited to single pairwise comparisons, introducing problems of multiple comparison corrections if the method is used to analyse further data. However, despite these drawbacks I have used it whenever possible.

A further problem arises in some experiments, where the data show complex 'patterns'. In these experiments, pairwise comparisons, or even ANOVA, are limited in their application. They cannot themselves tell us whether the pattern interpreted by the experimenter is the pattern indicated by the data – only whether or not the pattern tested by the experimenter is 'real'. This is discussed in Appendix F, where I run through the challenges that such an analysis presents.

Stimuli

Generating polychromatic textures

Texture perception is a complex discipline, and the cardinal dimensions of our perceptual visual 'texture-space' have yet to be unambiguously determined, though several attempts have been made (Rao and Lohse 1993; Gurnsey and Fleet 2001). One of the barriers to progress in the field is the difficulty in creating realistic textures that can be used to study it. Several texture-databases are available, however these are often limited in important respects. As an example, the widely-used Brodatz texture library was originally intended for 'artists and designers' (Brodatz 1966) and it is therefore no criticism to point out that it was created in the days before digital cameras were available, and the film-camera used to generate could not have been well characterised. This database is also achromatic, and seemed to be unavailable in a digital format when I undertook this research. Whilst fit for its original purpose, it is not suitable as a source of textures for my experiments.

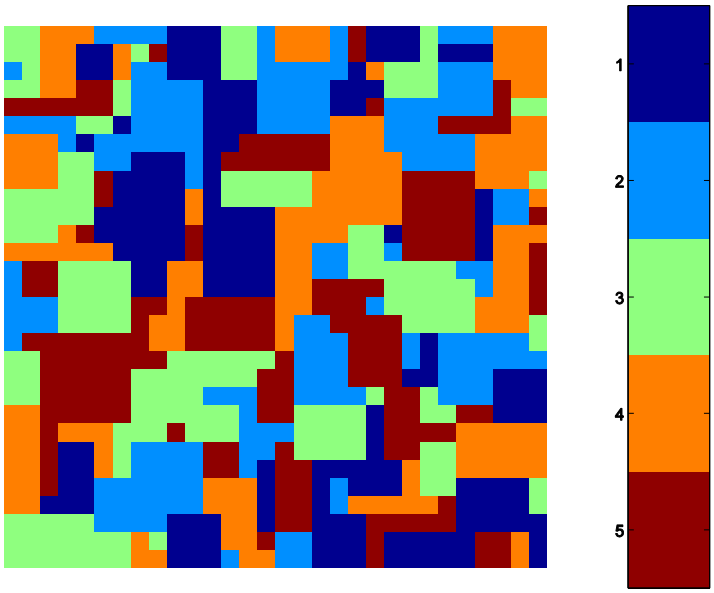


Figure 27: an example texture containing five component colours. The colourmap shown is for illustration purposes only, and the colours used in the experiment bore no relation to the colours shown in this figure.

Textures may also be relatively difficult to manipulate convincingly – for example if it is necessary to tile an irregular texture to produce a sufficiently large sample to entirely cover a surface without leaving visible boundaries. The most common methods of creating synthetic textures for use as visual stimuli include filtering random noise until it has a desired frequency spectrum, or using two-dimensional hidden-Markov models. The first method is strictly limited in its ability to recreate the richness of naturally occurring textures, and the second method is moderately complex to implement and produces stimuli whose properties may be difficult to control. I wished to investigate simultaneous induction in polychromatic surfaces, so it was unlikely to be easy to generate suitable textures using either of these techniques.

Although the use of naturalistic stimuli in experiments is always appealing, I felt that deficiencies in the texture databases available at the time would inevitably make it difficult to simulate natural stimuli with fidelity. In view of these problems, I decided to generate my own variegated textures in order to retain maximum control over their properties.

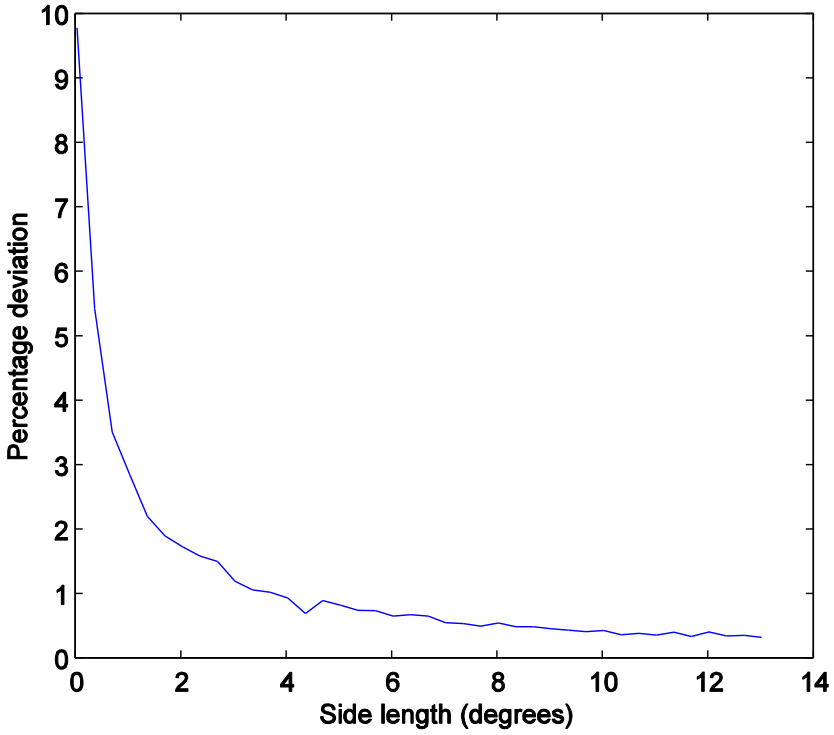


Figure 28: Variation in the chromatic content for randomly chosen texture patches of different sizes. The y-axis plots the mean difference between the proportions of squares coloured in each chromaticity, and the predicted proportions.

Figure 27 illustrates how one such texture was generated. A Matlab program was used to generate 'mini-Mondrian' stimuli by scattering coloured rectangles with pseudorandom dimensions onto a blank 'canvas'. Recently positioned rectangles occlude older rectangles, and at each iteration the area of the canvas that is covered by each colour is calculated. The colour of each new rectangle is chosen by determining the colour that is currently most lacking in the display, and the program stops when the areas of squares of each colour are equal to within an arbitrary limit. This limit depends on the size of the canvas, but for a small Mondrian target with five component colours, it is just over 0.5%. For larger canvas sizes such as Mondrian backgrounds, the imbalances are even less significant. Textures of arbitrary size can be created, and their properties are identical if they are rotated through 90^{o} or if the colours are interchanged.

A limitation to this algorithm is that it does not try to average the chromatic distribution on a local level, so small-scale deviations from the mean colour are possible though they are limited by the pseudo-random nature of the stimuli. This means that a patch chosen at random from the texture may contain a surfeit of some colours at the expense of the others. Figure 28 plots colour-biases against stimulus size, and shows that such random deviations are only significant in smaller patches. The chromatic distribution of Mondrian center patches is accurate as they are created independently from the background, but the chromatic distribution of the parts of Mondrian backgrounds immediately adjacent to target patches is random. Even in the worst case, the inhomogeneity is unlikely to affect any of the results reported here.

Apparatus

These experiments were all carried out using a 20” SGI monitor (curved screen). Automated characterisation was performed using a Matlab program to control the Minolta CS-100 Chromameter. To promote stable adaptation, observers viewed the computer monitor through a viewing box with white, diffusing walls. A chin-rest integral to the viewing box fixed the viewing distance at 57 cm.

The stimuli required for each experiment were generated in the form of uncompressed 24-bit colour 'bitmap' images prior to running each experiment. Stimulus presentation was performed under Windows 98 and XP using 'Presentation' – a software package used to present psychophysics experiments with accurate timings, to record observer responses by means of a gamepad, and to save this data to disk following each experimental session.

Analysis of the experimental results was performed using 'Matlab' and the 'psychofit' curve-fitting toolbox, now called 'psignifit'. More precise details are given for each experiment where required.

Experimental logistics

Observers for this series of experiments consisted of the two experimenters (KW and ACH), YL – then a fellow PhD candidate, and two final-year students OC and JW who undertook a limited amount of observing as part of their placements in the laboratory.

As in experiment 1, before starting to collect experimental data, new observers learned to perform the task by practicing the easiest task – judging the colours of stimuli presented against neutral, uniform-coloured backgrounds. As they became adept at this, they practiced with neutral articulated backgrounds and finally moved on to uniform and articulated colour-shifted backgrounds where the colour judgments are more difficult to make. The judgment of when they were ready to start collecting data was made arbitrarily based on the consistency of their data and their lapse rates. A common reason for a ‘lapse’ was to knowingly hit the wrong button due to unfamiliarity with the controls, and this improved with practice.

For each condition, at least two separate blocks of trials had to be run: one for the neutral trials, and one for the shifted trials. These could not be mixed at random, as this would cause changes in the observer's adaptational state over the duration of the experiment. For the same reason, I did not mix articulated and unarticulated targets and backgrounds. During any single block of trials, the only parameter to change from trial to trial was the mean chromaticity of the target square.

Over time, the 'neutral point' chosen by an observer may move along the inducing axis. Although for an individual observer these shifts are not necessarily large, it was still desirable to avoid any influence this may have had over the results. For this reason, conditions that are compared within a single graph were typically run in close temporal succession. Blocks measuring different conditions were mixed within a single observing session, and we attempted to collect full datasets as quickly as possible.

The exception to this rule was for some of the basic conditions, such as a uniformly coloured target viewed against a uniform background. I, observer KW, was able to devote more time to observing than the other observers and I always collected a fresh set of data for each experiment. But the other observers – all unpaid – were limited in their availability for the experiments. For this reason, we were reluctant to repeat data collection when it seemed very unlikely that doing so would make a material difference to our results. As an example, take figure 43. For observers YL and OC, the rightmost data point was taken from previous data, but all other datapoints were new measurements. Given that the data follows a consistent pattern it seems unlikely that repeating these measurements would have changed the data greatly.

The number of trials for each datapoint is quite variable. Again, I collected as many datapoints as possible for the full range of conditions. For the other observers we reduced the number of datapoints in order to reduce the observing time as much as possible, whilst simultaneously maintaining sufficiently small error bars for the null point. The effects measured in this experiment are far larger than those measured in experiment 1, so by comparison we needed relatively small numbers of trials in order to obtain robust patterns of results.

Experiment 2.1.1

Aims and predictions

The aim of this experiment was to investigate the effect of adding texture to the inducing background. Our prediction is, that by adding chromatic texture to the background we signify that the two objects are separable. Their ability to influence each other's colour should therefore be reduced. The strength of induction should be lessened when there is a textured background, compared to a uniformly coloured surface of the same mean chromaticity and luminance.

Methods

Stimuli

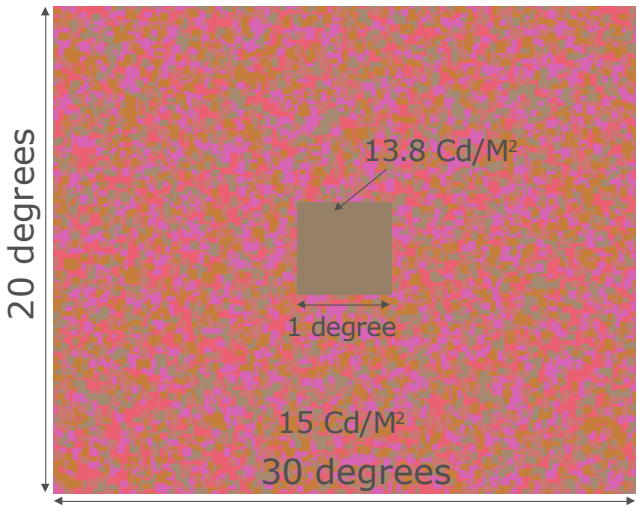


Figure 29: An example of an uniform patch set against a textured background, illustrating stimulus parameters including size and luminance. This diagram is not to scale.

The method shown in figure 25 was used to determine the strength of simultaneous chromatic induction for stimuli with an uniform background (not shown) or a textured 'Mondrian' background whose component chromaticities had +L-M contrasts of +/- 0.1, +/- 0.05 and 0 respectively. The optional mask was used in these experiments.

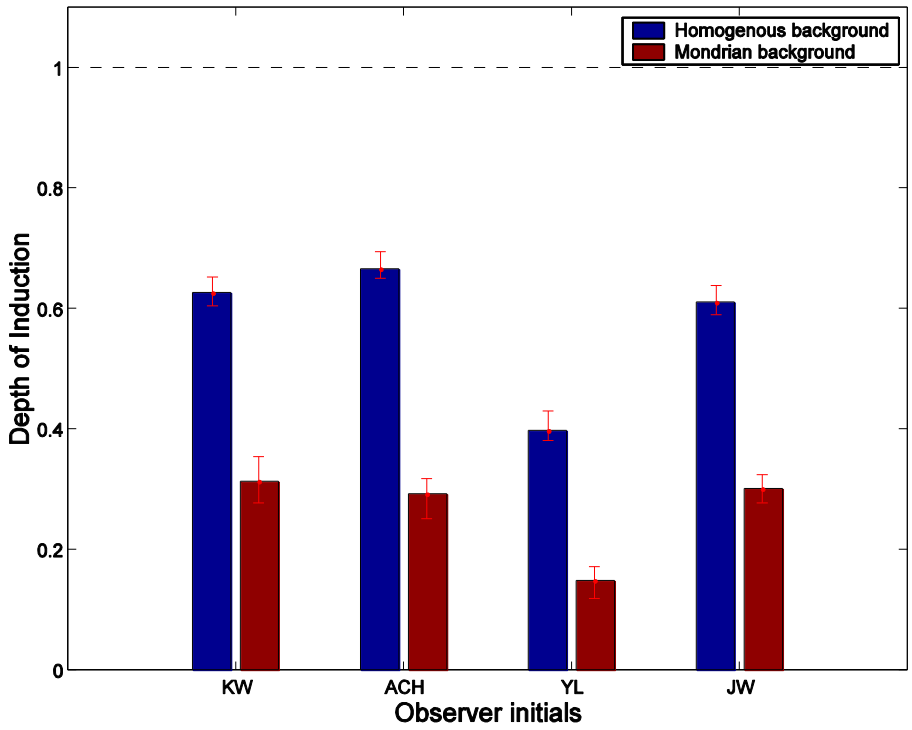


Figure 30: The strength of simultaneous contrast induction compared for the uniform background condition (filled bars) and the textured 'Mondrian' background (empty bars). Error bars indicate 95% confidence intervals for the threshold, as reported by the psignifit curve-fitting software. For all observers, texture in the background reduced the strength of simultaneous contrast induction.

Figure 30 shows our findings in this experiment. For all observers, the strength of contrast induction is approximately halved if the background contains an L-M texture with parameters similar to those described above. Our prediction is confirmed.

Table 4 provides further details of the analyses performed. P-values obtained using 'pfcmp' were highly statistically significant for all observers. Pseudo-'d' statistics, calculated using the variance of experiment 3.1, also indicate that the difference is highly functionally significant, being far larger than a typical sensory threshold.

Observer: ACH

Uniform Background Mean: 0.67 95% Confidence-interval 0.65 to 0.69 Mean number of trials per point: 20.0 (Max = 20; Min = 20)
Mondrian Background Mean: 0.29 95% Confidence-interval 0.26 to 0.32 Mean number of trials per point: 20.0 (Max = 20; Min = 20)
Difference size: 0.373
p-value: 0.0000
Cohen's 'd' statistic: 4.980

Observer: JW

Uniform Background Mean: 0.61 95% Confidence-interval 0.58 to 0.63 Mean number of trials per point: 20.0 (Max = 20; Min = 20)
Mondrian Background Mean: 0.30 95% Confidence-interval 0.28 to 0.32 Mean number of trials per point: 40.0 (Max = 40; Min = 40)
Difference size: 0.309
p-value: 0.0000

Observer: KW

Uniform Background Mean: 0.63 95% Confidence-interval 0.60 to 0.64 Mean number of trials per point: 53.1 (Max = 55; Min = 51)
Mondrian Background Mean: 0.31 95% Confidence-interval 0.29 to 0.34 Mean number of trials per point: 20.0 (Max = 20; Min = 20)
Difference size: 0.313
p-value: 0.0000
Cohen's 'd' statistic: 4.469

Observer: YL

Uniform Background Mean: 0.40 95% Confidence-interval 0.36 to 0.44 Mean number of trials per point: 20.0 (Max = 20; Min = 20)
Mondrian Background Mean: 0.15 95% Confidence-interval 0.11 to 0.17 Mean number of trials per point: 20.0 (Max = 20; Min = 20)
Difference size: 0.249
p-value: 0.0000
Cohen's 'd' statistic: 4.701

Table 4: Raw data relating to figure 30

Experiment 2.1.2

Aims and predictions

Here we investigate whether adding chromatic texture to target surfaces increases or reduces the perceived colour change that may be induced by varying the immediate background.

Here, we may attempt to make a more quantitative prediction. The cue of ensuring that background and target contain different textures should be equally valid if the background is textured, but the target is uniform, or if the background is uniform and the target is textured. For this reason, we should expect the reduction in the simultaneous contrast effect to be approximately identical in magnitude, and we will expect to find values between 0.25 and 0.35.

Method

As in experiment 2.1.1, the unmodified 2AFC method was used to compare values for the strength of simultaneous contrast induction for textured and uniform target squares, examples of which are shown in figure 31.

Stimuli

Figure 31: A series of 'Mondrian' target squares with the same internal contrasts, but varying along the L-M axis in terms of mean contrast relative to neutral. The 'reference' square has a contrast of 0; the background shift was along the same contrast, but had a total cone contrast of 0.1. The square contains five colours – red, green (LM contrasts of +/-0.1 respectively) blue and yellow (S-cone-contrasts of +/-0.7 respectively) and grey (neutral).

As can be seen, in this experiment we used targets that varied both along the +L-M axis, and along the S-axis. This was partly to facilitate some control experiments that we wished to carry out, where we combined the S-contrasts and L-M contrasts of the stimulus elements differently, to produce target chips with the same space-averaged luminances and chromaticities, but different sets of component colours.

Results

Adding texture to the test object dramatically inhibits contrast induction for all observers. This is consistent with our predictions, however induction appears considerably weaker than it was in the textured surround condition of the previous experiment.

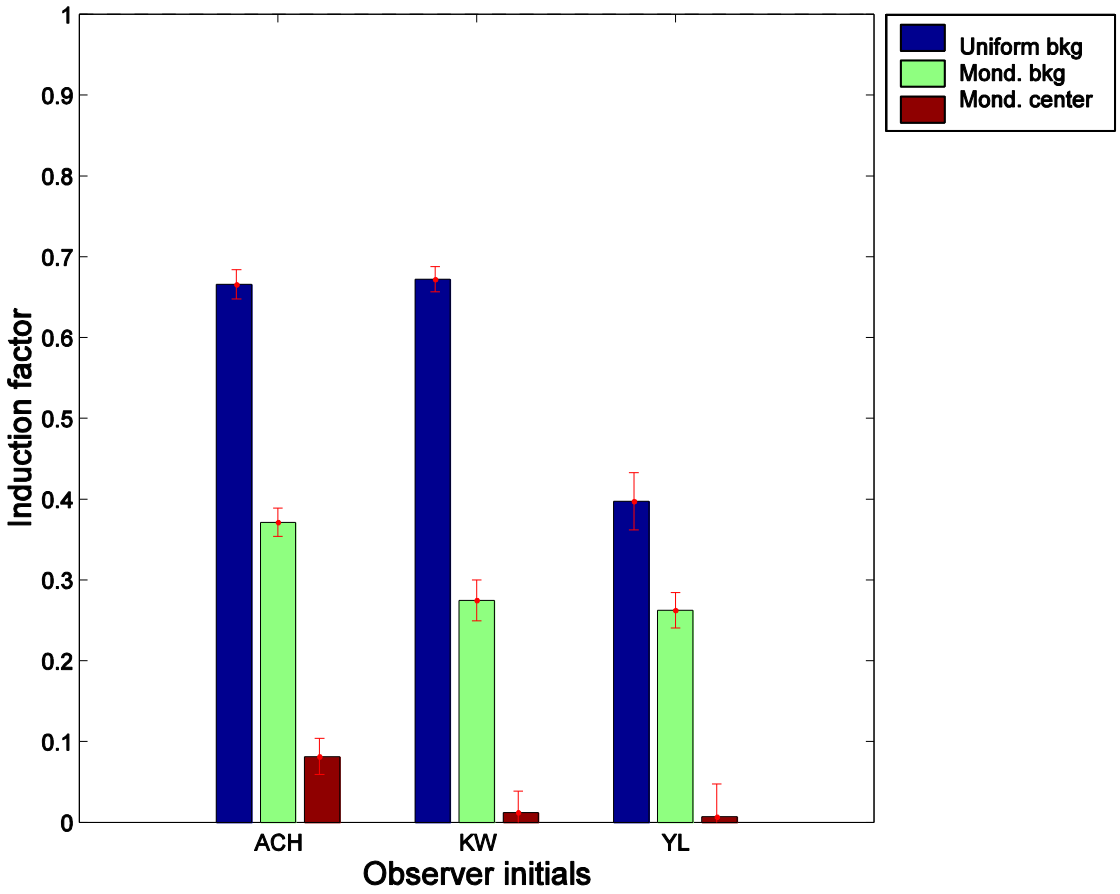


Figure 32: Induction when both target and background are uniformly textured, when only the background has a Mondrian texture, and when only the target has a Mondrian texture. Error bars indicate 95% confidence-intervals, as determined using 'psignifit'. Pairwise comparisons made with pfcmp show that all intra-observer differences are statistically significant at better than the 95% level.

Figure 32 shows that adding texture to the test object dramatically inhibits contrast induction for all observers, relative to the 'baseline' condition where induction occurs between untextured, uniform surfaces. This is qualitatively consistent with our predictions, however when plotted next to the condition for induction by a Mondrian-textured background, a marked asymmetry becomes apparent. Induction is weakened both when there is texture in the surround, and when there is texture in the target square, but it is inhibited far more strongly by texture in the target than texture in the surround. This is not a pattern that I would have predicted: the texture contrast indicating that the surfaces are discontinuous, should be an equally strong cue in both situations.

Table 5 shows the source data used to draw figure 32, including a pseudo-d statistic, calculated as previously described using standard deviation measures from experiment 3, comparing the strength of induction for the neutral targets and the strength of induction in a Mondrian target. These values are the largest reported in this thesis: larger even than those reported in table 4 for the comparisons between a uniform and Mondrian background.

Observer: ACH

Uniform target Mean: 0.67 95% Confidence-interval 0.65 to 0.69 Mean number of trials per point: 20.0 (Max = 20; Min = 20)

Mondrian target Mean: 0.08 95% Confidence-interval 0.006 to 0.10 Mean number of trials per point: 50.0 (Max = 50; Min = 50)

Cohen's Effect Size 'd': 11.024

p-values = 0 (uniform target/bkg v. Mond. Bkg; Mond. Bkg v. Mond. Target)

Observer: KW

Uniform target Mean: 0.63 95% Confidence-interval 0.60 to 0.64 Mean number of trials per point: 53.1 (Max = 55; Min = 51)

Mondrian target Mean: 0.01 95% Confidence-interval -0.01 to 0.03 Mean number of trials per point: 50.0 (Max = 50; Min = 50)

Cohen's Effect Size 'd': 9.441

p-values = 0 (uniform target/bkg v. Mond. Bkg; Mond. Bkg v. Mond. Target)

Observer: YL

Uniform target Mean: 0.40 95% Confidence-interval 0.36 to 0.44 Mean number of trials per point: 20.0 (Max = 20; Min = 20)

Mondrian target Mean: 0.00 95% Confidence-interval -0.04 to 0.04 Mean number of trials per point: 20.0 (Max = 20; Min = 20)

Cohen's Effect Size 'd': 7.484

p-values = 0 (uniform target/bkg v. Mond. Bkg; Mond. Bkg v. Mond. Target)

Table 5: Data used to draw figure 32

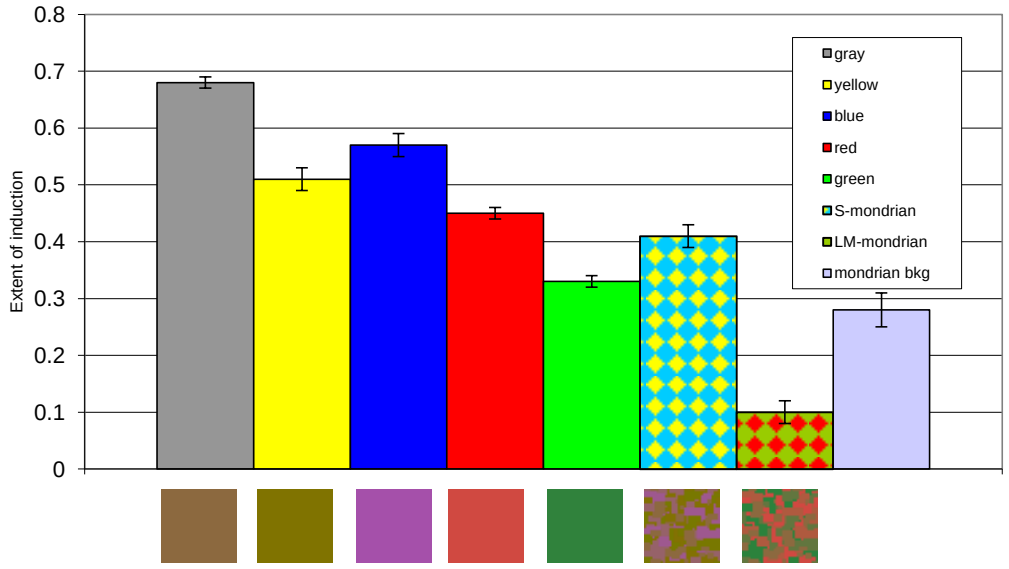


Figure 33: Control data for a single observer (KW) showing that induction is weaker for an L-M texture (value = 0.01) than for any of its individual component colours.

It is reasonable to question whether these results truly reflect a reduction in contrast induction due to the texture, or whether the reduction was merely due to the fact that the textured surfaces contain different colours from the uniform, gray test target. Figure 33 shows a control experiment carried out by observer KW, comparing values for induction in uniform target squares with different mean chromaticities. Induction is weaker for +L (red) and +M (green) stimuli, but it is still three to four times stronger than for textured stimuli. Moreover, only 2/5 of the target square is filled with such strong colours as were used in this control experiment.

In two additional sets of control experiments we showed that for two observers (YL, who was naïve and experimenter KW) it is possible to make good judgments about the mean colour of an object, despite it being composed of many constituent colours. In the first set of control experiments, we swapped the +/-LM and +/-S contrasts of the various component colours between the reference and test targets so that there was no correspondence between the colours of the stimuli presented to the observers in each case, although the mean contrasts between the different stimulus elements were the same in each case. In the second set of control experiments, we also swapped the positions of the elements at random, so that there was no correspondence between the texture on the reference and test targets. In both cases we obtained the same results: induction was always very weak for textured targets, whatever their individual component colours or element distributions.

Condition	***Mean induction factor***	***95% confidence interval***	***Number of trials per point (mean)***	***Min. # of trials***	***Max. # of trials***
Gray	0.63	0.6 - 0.64	53.12	51	55
Yellow	0.52	0.53 - 0.55	50	50	50
Blue	0.57	0.55 - 0.59	50	50	50
Red	0.39	0.38 - 0.4	50	50	50
Green	0.33	0.32 - 0.35	50	50	50
S-isolating Mondrian	0.41	0.39 - 0.42	50	50	50
Achromatic Mondrian	0.24	0.21 - 0.30	20	20	20
LM-isolating Mondrian	0.04	0.02 - 0.06	50	50	50
Mondrian background	0.27	0.25 - 0.31	50	50	50

Table 6: data used to draw figure 33. Inferential statistical tests were not performed on these data

Experiment 2.2.1

Aims and predictions

In this experiment, I examine whether a textured background is able to induce a perceived colour change in a textured target. I also quantitatively investigate how varying the internal chromatic contrast of the surround texture influences induction in both textured and uniform centers. The results of these experiments will help address a number of questions relating to why simultaneous contrast induction appears to be weakened by the presence of texture in objects.

Before carrying out this experiment, I had a number of working theories about why texture in the surface or in the target inhibits simultaneous contrast induction:

1) **Segmentation of the target from the surround:** A texture difference between the target and background is a strong segmentation cue that suggests that they may be separate from one another. Any demonstration of simultaneous contrast must use two or more surfaces and it is possible that the brain modulates simultaneous contrast induction based on whether or not surfaces are 'segmentable'. In this case, filling both the surround and the target surface with similar textures should reduce cues to segmentation, relative to the case where either surround or target is textured. If this is the case, we would predict that simultaneous contrast induction should be restored to the same levels as were measured for the uniform target/surround condition.

2) **Reinforcement of local colour perception due to polychromaticity:** Surfaces that contain multiple chromaticities provide more information about their illumination than do uniformly coloured surfaces. As such it may be theoretically possible to attain colour-constancy for a polychromatic surface viewed in the void condition – without reference to other objects. If this is the case, polychromatic surfaces may plausibly be protected against chromatic shifts induced by surrounding surfaces, and replacing a uniform background with a textured one should not reinforce simultaneous contrast induction. In this case I would expect chromatic induction to be roughly equally strong when both target and surround are textured, as when the target alone is textured.

Boundary effects: Some models of colour perception involve edge detectors working in conjunction with 'filling in' mechanisms. The 'edge detector' cells determine the colour contrast of enclosed borders and work together to determine the colour of a

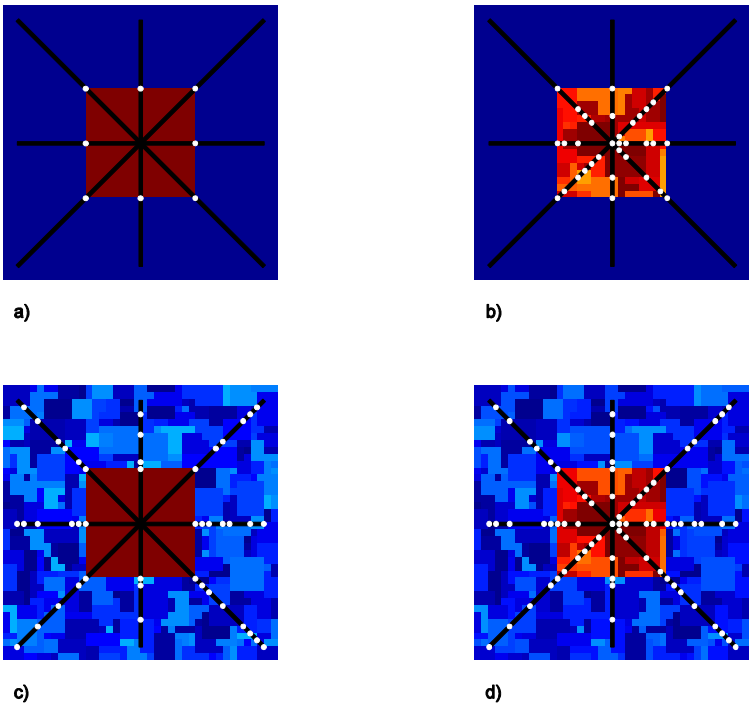


Figure 34: When a 'mondrian' texture is added to either the target or the surround, chromatic boundaries are created.

a) shows the 'standard' stimulus used to elicit simultaneous contrast induction: an uniformly coloured target against an uniformly coloured surround. The only chromatic border present in the image is that between the target and surround.

b) shows a condition with a textured target. Here, distant parts of the surround may be connected to the external border of the target by paths that do not cross any chromatic borders. However, they may not be connected to points far inside the target surface without crossing any chromatic borders.

c) shows a condition with a textured surround, and an uniform target. Those parts of the surround in immediate proximity to the target's may be connected to the center of the target, with only a single border created.

d) shows a condition where both surround and target are textured. It is impossible to connect any two distant points, without having to cross a number of chromatic borders.

delineated surface; the 'filling-in' mechanisms are said to propagate the colour determined by the edge detectors across the surface of the object, causing it to appear an uniform colour. These models of colour perception are supported by both psychophysical and neurophysiological evidence. Figure 34 and figure 35 demonstrate such 'filling-in' effects.

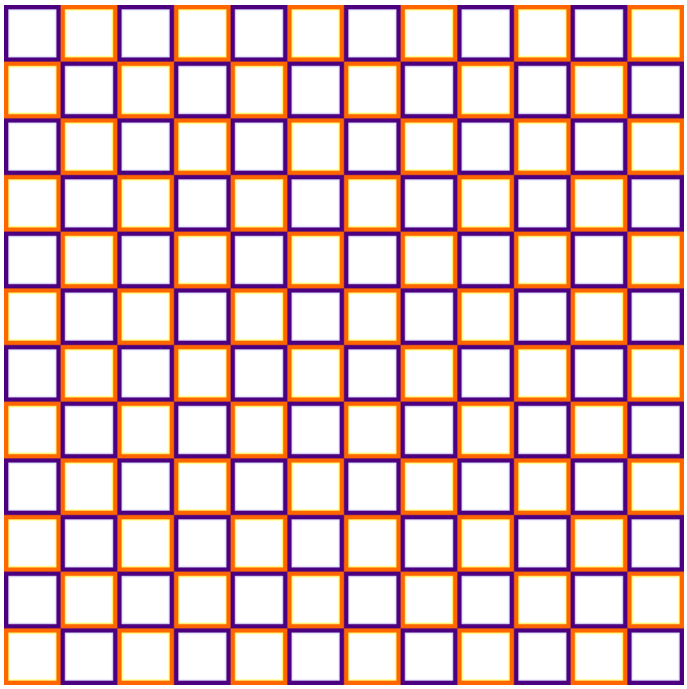


Figure 35: 'Filling-in'. This checkerboard figure is composed of dark purple and light orange/yellow squares with white centers. Most people judge that the white squares with orange borders appear slightly orange, and that the purple-bordered squares are slightly purple in appearance.

Various areas of the brain are currently believed to play an important role in chromatic "filling-in". Different populations of cells have been identified whose functions seems to be compatible either with the 'edge-detector' role, or a in 'filling-in' the target centers based on information provided by the 'edge detector' cells.

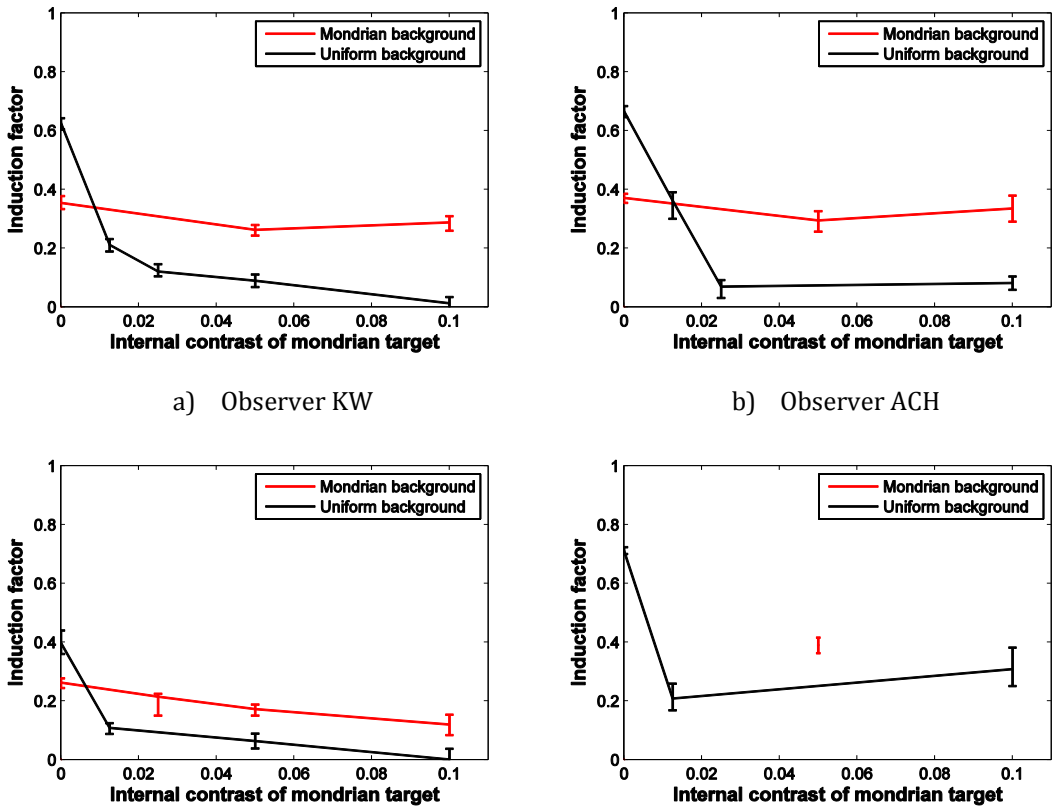
In textured surfaces such as the 'mini-Mondrians' used in my stimuli, any path linking two well-spaced points will necessarily cross a number of chromatic borders. If the 'filling-in' process occurs separately for each chromatic border then we might expect induction effects from the background to decay more rapidly with distance. 'Filling-in' is a common phenomenon that can work for texture, form, colour and other visual attributes also.

Retinex models can be applied to 'Mondrian' scenes that are large-scale models of my Mondrian patches, to provide an estimate of the illuminant, and they do this by a process of measuring edge-contrasts and 'filling-in'. In this sense, my hypothesis that boundary-effects may be responsible for stabilising the appearance of a textured surface against chromatic contrast induction may be related to hypothesis 2: that colour-constancy mechanisms stabilise the colour appearance of the textured target by providing sufficient information for the visual system to make an estimate of the target object's illumination without needing to take into account remote surfaces that may vary from setting to setting, and therefore provide less dependable cues to colour constancy.

If boundary effects are responsible for the reduction in induced contrast in textured targets, or for the fact that texture in the backgrounds can weaken contrast effects, then adding texture to both target and surround may be expected to reduce contrast effects more than adding texture to either target or surround alone. In practice, adding texture to the target already inhibits simultaneous contrast induction strongly enough that it could be difficult to detect a further reduction in the strength of simultaneous contrast induction, when both target and surround are textured.

Methods and stimuli

As in experiments 2.1.1.-2.1.2, the unmodified 2AFC method was used to compare values for the strength of simultaneous contrast induction for textured and uniform target squares, examples of which are shown in figure 35. The level of internal contrast within the Mondrian-textured target was varied, and the strength of contrast induction was measured, both for uniform and Mondrian backgrounds. In the Mondrian background condition, the internal L-M contrast of the texture was always ±0.1. The stimulus dimensions and paradigm were identical to those used in experiments 2.1.1.-2.1.2.



a) Observer KW

b) Observer ACH

c) Observer YL

d) Observer OC

Figure 36: induction in textured target squares by a mondrian background (black line) or an uniform background (red line). For all four observers, the red line is higher than the black line for at least part of the course, indicating that a textured background can partially restore the simultaneous contrast effect for textured targets.

Figure 36 shows that the strength of the contrast effect is restored to approximately the same level as was measured when only the background was textured and the target was uniform, showing that the reduction in contrast strength must be due in part to differences between target and surround and not to the increased number of chromatic transitions in the stimuli (contrast is known to be inhibited by chromatic borders (Zaidi, Yoshimi et al. 1991)). It also appears to rule out the hypothesis that colour-constancy effects at a very local level were able to reinforce the appearance of the target against perturbations from the surround, because all the information necessary for local colour-constancy effects to act will still have been present in the Mondrian target, despite the additional texture in the Mondrian background.

a) Observer: ACH

Uniform backgrounds; textured target

Contrast level: 0.00000 Mean: 0.666 95% Confidence-interval 0.646 to 0.683 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.01250 Mean: 0.362 95% Confidence-interval 0.300 to 0.390 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.02500 Mean: 0.069 95% Confidence-interval 0.031 to 0.091 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.10000 Mean: 0.081 95% Confidence-interval 0.059 to 0.103 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Mondrian backgrounds; textured target

Contrast level: 0.00000 Mean: 0.371 95% Confidence-interval 0.354 to 0.385 Mean number of trials per point: 30.00 (Max = 30; Min = 30)

Contrast level: 0.05000 Mean: 0.294 95% Confidence-interval 0.256 to 0.326 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.10000 Mean: 0.335 95% Confidence-interval 0.290 to 0.379 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level 0: p value = 0.000

Contrast level 0.1: p value = 0.000

b) Observer: KW

Uniform backgrounds; textured target

Contrast level: 0.00000 Mean: 0.626 95% Confidence-interval 0.604 to 0.642 Mean number of trials per point: 53.13 (Max = 55; Min = 51)

Contrast level: 0.01250 Mean: 0.211 95% Confidence-interval 0.189 to 0.231 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.02500 Mean: 0.120 95% Confidence-interval 0.104 to 0.145 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.05000 Mean: 0.089 95% Confidence-interval 0.068 to 0.110 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.10000 Mean: 0.012 95% Confidence-interval -0.010 to 0.034 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Mondrian backgrounds; textured target

Contrast level: 0.00000 Mean: 0.354 95% Confidence-interval 0.333 to 0.377 Mean number of trials per point: 75.00 (Max = 100; Min = 50)

Contrast level: 0.05000 Mean: 0.263 95% Confidence-interval 0.243 to 0.279 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.10000 Mean: 0.287 95% Confidence-interval 0.260 to 0.309 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level 0: p value = 0.000

Contrast level 0.05: p value = 0.000

Contrast level 0.1: p value = 0.000

C) Observer: OC

Uniform backgrounds; textured target

Contrast level: 0.00000 Mean: 0.712 95% Confidence-interval 0.700 to 0.723 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.01250 Mean: 0.209 95% Confidence-interval 0.168 to 0.259 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.10000 Mean: 0.308 95% Confidence-interval 0.251 to 0.381 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Mondrian backgrounds; textured target

Contrast level: 0.05000 Mean: 0.390 95% Confidence-interval 0.363 to 0.415 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

d) Observer: YL

Uniform backgrounds; textured target

Contrast level: 0.00000 Mean: 0.397 95% Confidence-interval 0.360 to 0.440 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.01250 Mean: 0.108 95% Confidence-interval 0.088 to 0.124 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.05000 Mean: 0.064 95% Confidence-interval 0.039 to 0.089 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.10000 Mean: 0.001 95% Confidence-interval -0.038 to 0.037 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Mondrian backgrounds; textured target

Contrast level: 0.00000 Mean: 0.262 95% Confidence-interval 0.244 to 0.277 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.02500 Mean: 0.215 95% Confidence-interval 0.150 to 0.224 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.05000 Mean: 0.172 95% Confidence-interval 0.150 to 0.188 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.10000 Mean: 0.120 95% Confidence-interval 0.084 to 0.153 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level 0: p value = 0.000

Contrast level 0.05: p value = 0.000

Contrast level 0.1: p value = 0.002

Table 7: The raw data used to plot figure 36a-d. Where possible, p-values without multiple-comparison corrections have been calculated for any pairwise comparisons that may be made.

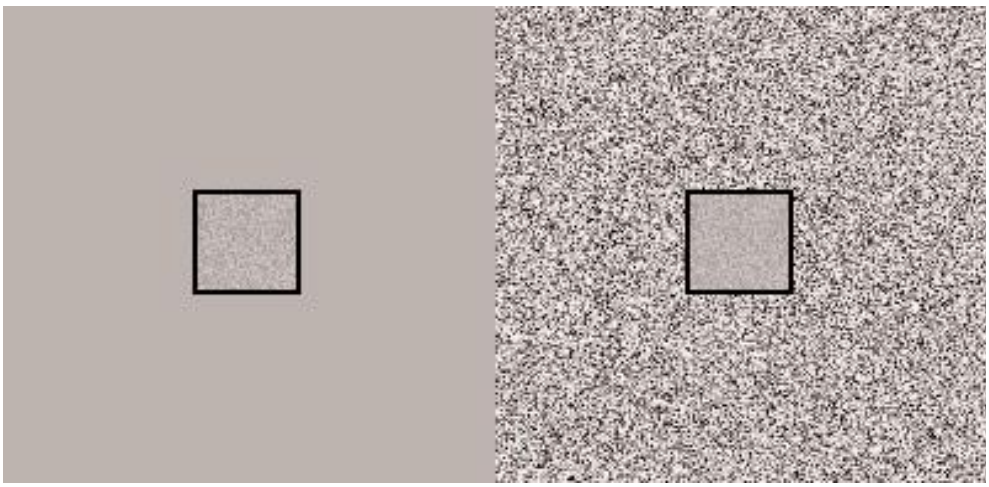


Figure 37: Texture-texture suppression after Chubb, Solomon & Sperling (1989). To most people the textured patch in the center of the left-hand square appears more contrasty than the patch on the right-hand side of the diagram, even though both are in fact identical.

Yet, several factors still cast doubt on this interpretation that the results from experiments 2.1 can be explained by texture-mediated segmentation. Firstly, we have still not explained why texture in the target seems so much more effective in inhibiting simultaneous induction than texture in the surround. Is there any reason why the segmentation cues should be stronger in one condition than the other? Secondly, our prediction was that a condition where both target and surround were textured should be equivalent to a condition where both target and surround were uniformly coloured. This is clearly not the case – when the stimuli are textured, simultaneous contrast induction is still weaker than if the stimuli are uniformly coloured.

An alternative explanation is that contrast-contrast suppression occurs – causing the texture in the target object to become less salient to the neural circuits responsible for simultaneous contrast that are normally inhibited by it. An example of achromatic texture-texture suppression is given in figure 37, which shows a uniform background next to a textured background with two identical textured target patches set in the centre of each. If suppression by the texture in the surround were sufficiently strong as to nullify the effect of the texture in the target's center, then from the point of view of the neural mechanism involved, the configuration where both surround and target were textured would be equivalent to the condition when only the surround was textured.

If this is true, this observation casts doubt on the suggestion that polychromaticity may be important in surface colour perception. Why would we have cells that perform filling in and edge-detection, if natural surfaces have so many chromatic borders that filling-in and edge detection are not viable

Experiment 2.2.2

Aims

To discover whether the effect that we have found is due to texture-segmentation or some other property of the stimuli, such as the chromatic boundaries that are present in all textured stimuli. Are the mean properties of the background or the properties in the region local to the target important in determining whether contrast-induction is suppressed?

Methods and stimuli

Despite its position in this chapter, this experiment was my first to investigate how variation in the local surrounds influences simultaneous contrast. As such, the stimuli differ from those used in the other experiments described in this chapter, which evolved from it. The stimuli here are smaller (14^o square) and were viewed through a black-walled box, rather than in a white-walled viewer. The backgrounds were more structured checker patterns and there was no 'mask' stage; however in other respects the experimental method used was similar to that used elsewhere in this chapter.

On each trial the following sequence was presented: prototypic stimulus (500 ms); test stimulus (500 ms); neutral background (minimum 500ms). No mask was used in this set of experiments. The test stimulus was identical to the first in its spatial configuration; only the colours differed. The background colours were uniformly translated along the (L-M) axis in the isoluminant plane relative to their prototypic colours (see figure 39), always with a constant total cone contrast shift of 0.1.

In all tasks, the central target's size is 1^o; the background size is 14^o x 14^o. The prototypic stimulus consisted of an achromatic central square (size 1^o; CIE coordinates x= 0.321, y= 0.337; luminance = 28 Cd/M^2) against a uniform or complex background (size 14^o x 14^o; mean chromaticity, CIE coordinates x= 0.321, y= 0.337; luminance 30 Cd/M^2).

For each background type, 5 different backgrounds were generated, in each of which the configuration of the chromatic elements was varied at random. On each trial, one of the 5 backgrounds was selected at random, but remained identical between the prototypic and test stimuli, except for the global (L-M) shift. This ensured that long-term adaptation to the background pattern could not occur, and reduced the risk that atypical arrangements of the background squares may influence the strength of simultaneous contrast, as measured for particular conditions.

Observers were instructed to maintain fixation on the central square throughout each trial. Three observers (myself, my supervisor, and one 21-yr-old male, JW, who was naïve to the purposes of the experiment), participated in the experiment. All of us had normal colour vision, as verified with the Farnsworth-Munsell 100-hue test.

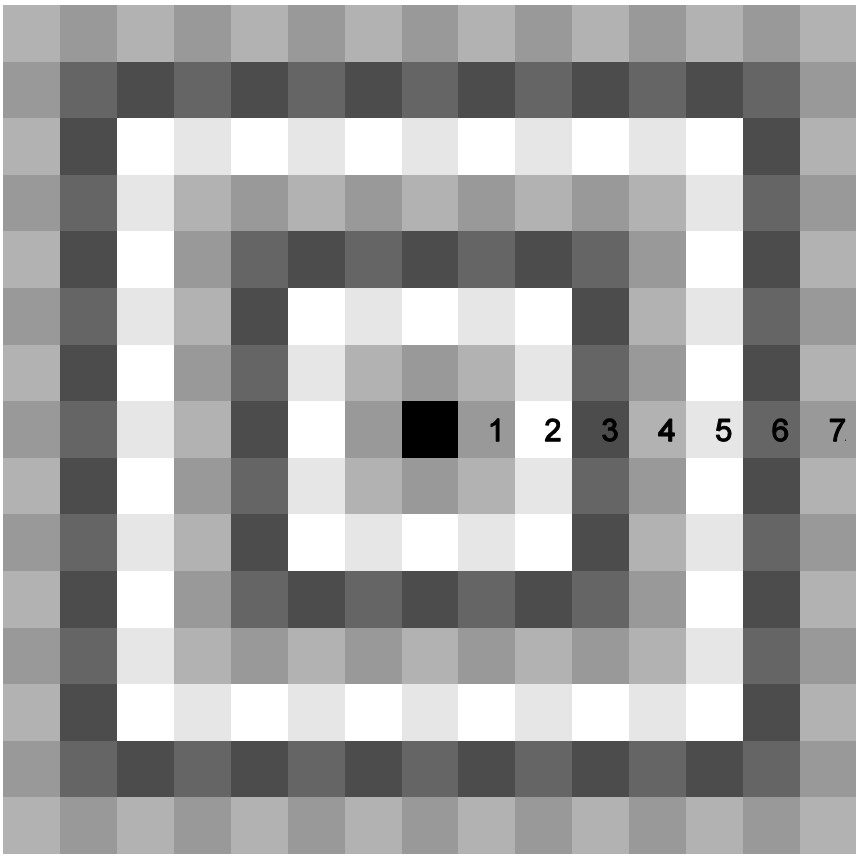


Figure 38: Preparing the background stimuli. Concentric rings of squares are arranged around a central square (black), that would normally be covered by the target square. The first ring (1) contains eight squares; the second (2) contains sixteen and further rings similarly contain multiples of eight squares.

Both the target stimuli and the backgrounds used in this experiment were pre-generated using Matlab, and were stored as 24 bit uncompressed bitmap images. The background images were composed of a number of concentric circles, each of which contained a multiple of 8 squares. These squares could be either 1 degree (the same size as the target square), 1/3 of a degree, 1/9 of a degree or 1/32 of a degree in size, allowing the spatial frequency statistics of the backgrounds to be varied. Backgrounds with higher spatial frequency components contained more concentric rings, so that the spatial extent of the background remained the same for all spatial frequency conditions.

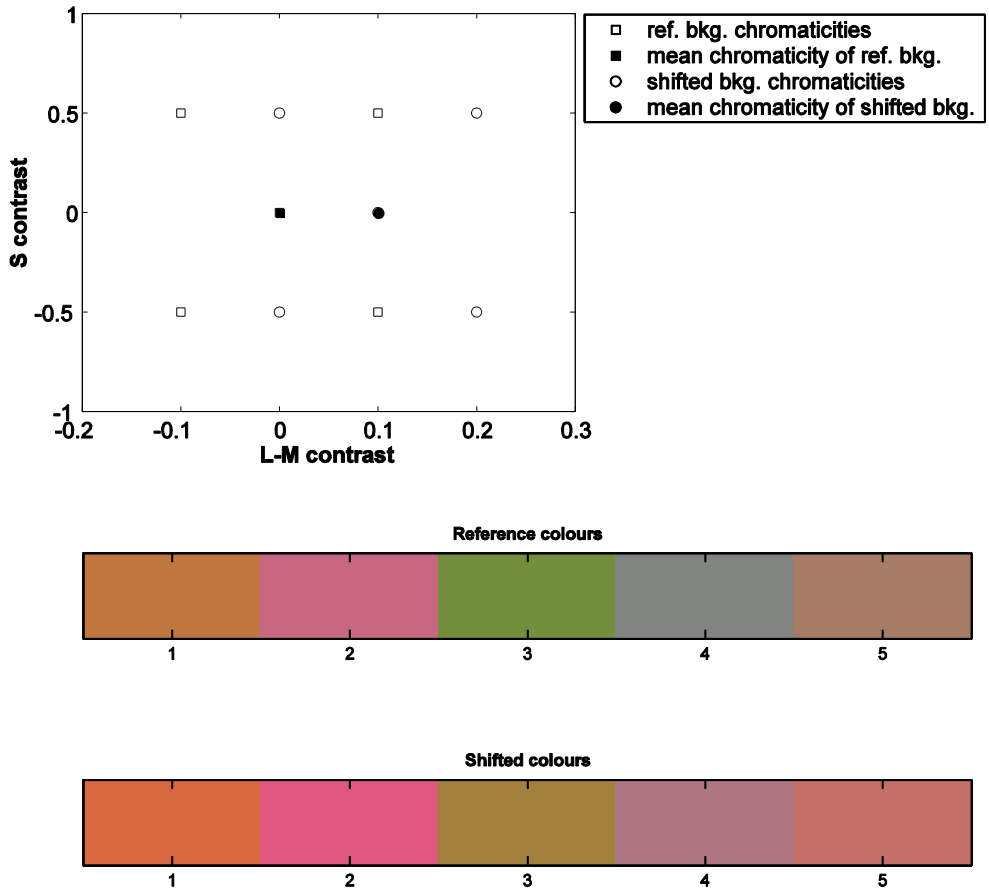


Figure 39: Background colours used in the experiment

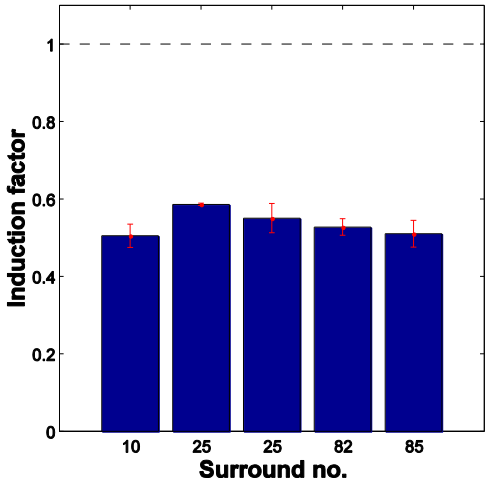
The top graph plots the reference and shifted background colours used in these experiments, in the modified McLeod-Boyton colourspace described in the methods section. Reference colours are plotted using square symbols; shifted colours using circular symbols. The space-averaged chromaticities for the reference and shifted backgrounds are plotted as filled figures.

Two colour-bars are plotted below the colourspace graph. The upper graph shows the reference colours, and the lower graph shows the shifted colours. The first four entries show the colours used in the background; the fifth colour is the space-averaged equivalent of the preceding colours.

Each ring was filled with equal numbers of four distinct colours, arranged at random. For example, there were eight squares in the first ring so each surround colour was represented twice; in the second ring there were sixteen squares therefore four of each colour, and so on. In this way the background took on a pseudo-random appearance, yet on a global level it was stabilised against deviations in colour from the mean. On a local level such deviations were still possible, as is shown by figure 40a that illustrates some of the 330 possible arrangements of the four colours in the first ring for the condition

when the background squares were the same size (1^{o}) as the target. For example in some conditions, all the L+ shifted squares (orange and magenta) may be adjacent, whereas in other conditions, L+ and L- squares may be interleaved.

a)



b)

Figure 40: Control experiment examining whether induction differs significantly for different arrangements of the eight immediate squares surrounding the target for the 1 square-per-degree backgrounds.

a) The 330 possible permutations of four distinct surround colours, ignoring equivalent figures that may be generated through rotation or reflection of these prototypes. In a control experiment, the strength of induction was compared for backgrounds 255, 85, 82, 102 and 250 that were chosen to represent distinct arrangements of the eight squares in the background.

For example, in background 255 all the L+ shifted parts of the stimuli (magenta and orange) are contiguous but in background 250 the magenta elements and the orange elements are on opposite sides of the target square, and are not contiguous. In background 82, no two adjacent squares areas are the same colour.

McCann (2003) showed that achromatic induction varied unpredictably as different numbers of white and black squares were placed in different arrangements around a

gray target square. My experiments differ in that the number of squares of each colour is always the same, yet they are arranged differently from background to background. Does the strength of induction vary significantly for different arrangements of the background squares?

```
Comparison: 10 v. 25: p-value = 0.00000
Comparison: 10 v. 82: p-value = 0.19372
Comparison: 10 v. 85: p-value = 0.14946
Comparison: 25 v. 82: p-value = 0.16106
Comparison: 25 v. 85: p-value = 0.08386
Comparison: 82 v. 85: p-value = 0.34958
Cohen's 'd' statistic: 0.12
```

Table 8: p values for all possible pairwise comparisons in figure 40. These statistics were calculated using Wichmann and Hill's 'psignifit' function. Only one is statistically significant, but most importantly the effect size is far smaller than many of those exhibited in other experiments reported in this chapter. No corrections have been made for multiple comparisons. The value for Cohen's 'd' statistic is a pseudo-value representing the difference between the largest and smallest levels of induction shown on the graph, relative to the standard deviation of KW's settings in experiment 3.1.1

```
Surround no: 10       Mean: 0.0505   95% Confidence-interval 0.047 to 0.054
Surround no: 25       Mean: 0.0586   95% Confidence-interval 0.059 to 0.059
Surround no: 82       Mean: 0.0550   95% Confidence-interval 0.051 to 0.058
Surround no: 85       Mean: 0.0527   95% Confidence-interval 0.051 to 0.056
```

Table 9: means and confidence intervals for the data shown in figure 40.

Observer KW carried out a control experiment, the results of which are shown in figure 40b and table 8. They show that induction can vary slightly between different surround arrangements, but the differences between induction measured for these configurations are smaller than those differences that are considered important later in the experiment, such as those shown in figure 30. The results discussed later suggest that the ring of stimuli closest to the target square have more influence than those further away. Secondly, as the number of randomly-placed elements increases it becomes less likely that their distribution will be strongly skewed towards an extreme. In view of these arguments it seems unlikely that the precise arrangement of the coloured-squares in noncontiguous parts of the background, or that the arrangement of the ring of squares contiguous to the target for the higher-spatial-frequency backgrounds – will strongly influence the results.

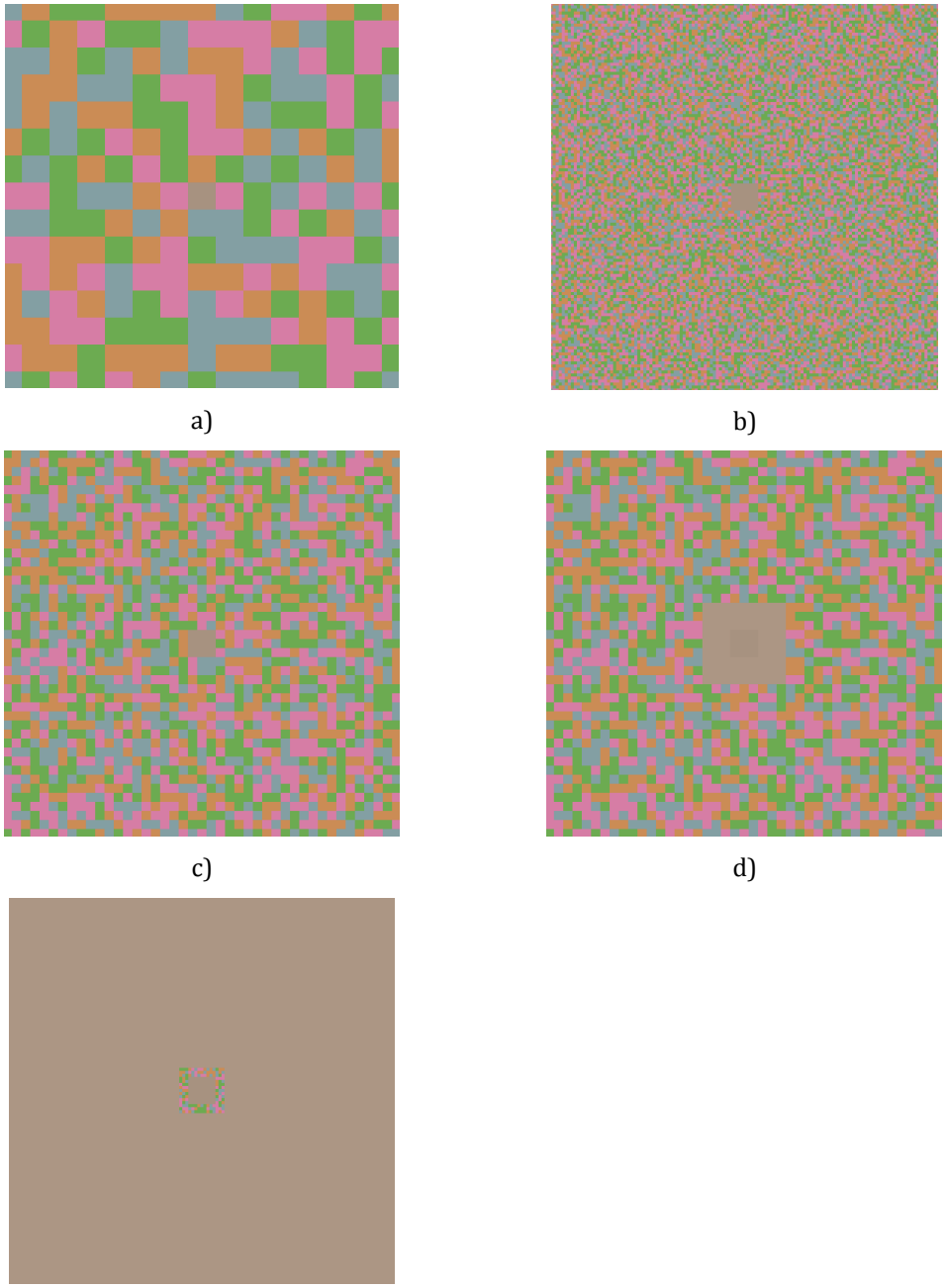


Figure 41: Examples of some of the stimuli used in the experiments

a) Background composed of 1^o squares

b) Background composed of $1/9^o$ squares

c) Background composed of $1/3^o$ squares

d) Background composed of $1/3^o$ squares with a 1^o wide uniform annulus

e) Uniform background with a $1/3^o$ wide modulated annulus with 9 squares per degree

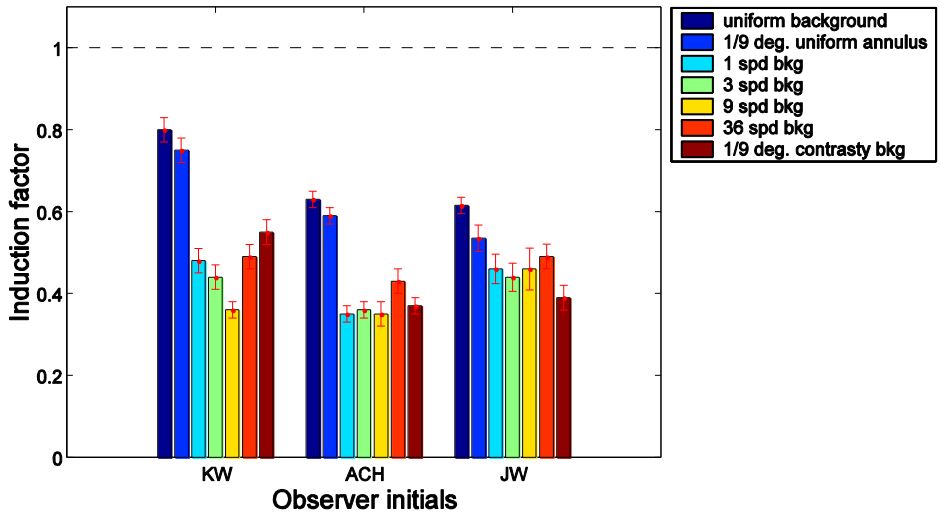


Figure 42: Induced contrast (measured as the total cone contrast of the central test square, relative to the prototypic square chromaticity, which appears neither 'red' nor 'green' against the L+-shifted background, as described in text), for different background conditions, for three observers. Error bars indicate 95% confidence-intervals. In the legend, 'spd' refers to 'squares per degree' and 'bkg' to 'background'. See figure 41 for further details.

Results

Our findings were broadly similar for all observers, and are shown in figure 42. The strongest measures for induction were obtained for a uniform target viewed against a uniform background, or for a uniform target separated from a uniform background by a $1/9^{o}$ uniform annulus. Induction is weakened whenever the local background is articulated – whether this articulation is global or confined to a $1/9^{o}$ annulus. The degree of weakening is independent of spatial frequency for observers ACH and JW for spatial frequencies up to 9 squares per degree, and only weakly dependent on spatial frequency for observer KW. However, induction is partially restored for all observers for the 36 SPD condition, a finding that may reflect our reduced sensitivity to such chromatic stimuli with high spatial frequencies.

Observer: KW

Condition: Uniform background Mean: 0.81 95% Confidence-interval: 0.79 to 0.84 Number of trials per point 20

Condition: 1/9 deg uniform ring Mean: 0.78 95% Confidence-interval: 0.74 to 0.81 Number of trials per point 20

Condition: 1 spd background Mean: 0.45 95% Confidence-interval: 0.41 to 0.48 Number of trials per point 20

Condition: 3 spd background Mean: 0.44 95% Confidence-interval: 0.41 to 0.47 Number of trials per point 20

Condition: 9 spd background Mean: 0.37 95% Confidence-interval: 0.33 to 0.41 Number of trials per point 20

Condition: 36 spd background Mean: 0.56 95% Confidence-interval: 0.52 to 0.59 Number of trials per point 20

Condition: 1/9 deg. art. bkg. Mean: 0.55 95% Confidence-interval: 0.50 to 0.58 Number of trials per point 20

Observer: ACH

Condition: Uniform background Mean: 0.63 95% Confidence-interval: 0.60 to 0.65 Number of trials per point 20

Condition: 1/9 deg uniform ring Mean: 0.59 95% Confidence-interval: 0.56 to 0.61 Number of trials per point 20

Condition: 1 spd background Mean: 0.35 95% Confidence-interval: 0.31 to 0.37 Number of trials per point 20

Condition: 3 spd background Mean: 0.36 95% Confidence-interval: 0.31 to 0.38 Number of trials per point 20

Condition: 9 spd background Mean: 0.31 95% Confidence-interval: 0.30 to NaN Number of trials per point 20

Condition: 36 spd background Mean: 0.43 95% Confidence-interval: 0.40 to 0.46 Number of trials per point 20

Condition: 1/9 deg. art. bkg. Mean: 0.37 95% Confidence-interval: 0.34 to 0.39 Number of trials per point 20

Observer: JW

Condition: Uniform background Mean: 0.56 95% Confidence-interval: 0.52 to 0.58 Number of trials per point 20

Condition: 1/9 deg uniform ring Mean: 0.53 95% Confidence-interval: 0.50 to 0.56 Number of trials per point 20

Condition: 1 spd background Mean: 0.48 95% Confidence-interval: 0.45 to 0.52 Number of trials per point 20

Condition: 3 spd background Mean: 0.44 95% Confidence-interval: 0.41 to 0.47 Number of trials per point 20

Condition: 9 spd background Mean: 0.46 95% Confidence-interval: 0.41 to 0.49 Number of trials per point 20

Condition: 36 spd background Mean: 0.49 95% Confidence-interval: 0.46 to 0.51 Number of trials per point 20

Condition: 1/9 deg. art. bkg. Mean: 0.40 95% Confidence-interval: 0.37 to 0.42 Number of trials per point 20

Table 10: details of the data used to draw figure 42. The abbreviation 'spd' refers to 'squares per degree'; 'arg. bkg.' refers to 'articulated background'

Observer: KW

Uniform background v. uniform annulus:	0.231
Uniform background v. 1 square per deg:	0.000
1 spd versus 36 spd:	0.023
Uniform background v. contrasty annulus:	0.000

Observer: ACH

Uniform background v. uniform annulus:	0.282
Uniform background v. 1 square per deg:	0.000
1 spd versus 36 spd:	0.000
Uniform background v. contrasty annulus:	0.000

Observer: JW

Uniform background v. uniform annulus:	0.425
Uniform background v. 1 square per deg:	0.073
1 spd versus 36 spd:	0.668
Uniform background v. contrasty annulus:	0.000

Table 11: p-values for selected pairwise comparisons, generated with 'pfcmp'. No multiple-comparison corrections have been made.

Conclusions

The object of this experiment was to determine whether textured surfaces affect contrast due to their global properties, or whether the effect of texture due to its action at boundaries between objects? The results appear to support the notion that local stimulus features are overwhelmingly important in modulating the strength of simultaneous contrast induction, suggesting that texture could modulate simultaneous contrast induction through boundary effects. However, if true these results still sit uncomfortably with the finding of experiment 2.2.1 that induction could be restored under some circumstances by manipulations that increase the number of chromatic boundaries in a stimulus (adding texture to the background, when previously only the target was textured).

Experiment 2.2.3

Aims

To discover whether texture suppression is likely to influence simultaneous contrast induction in the 'real-world' I investigate the relationship between the extent of contrast induction and the internal contrast of the texture in the target square.

Methods and Stimuli

In this experiment, I investigated only L-M background shifts, but I examined the effect of varying both the internal contrast of the texture in the target square or the chromatic axis of its texture. I used squares textured along all three cardinal axes in colourspace, and additionally investigated induction by backgrounds that varied in terms of their internal L-M texture contrast, giving a total of four different conditions:

a) +L-M texture (the same axis as the background shift)

b) S-cone texture

c) Achromatic texture

d) Varying the contrast of the background texture

These four conditions were investigated using stimuli similar to those in experiment 2.1.1, with a mask. The total-cone-contrasts of the achromatic texture and the L-M texture were made equal to aid comparisons. S-cone contrasts are less perceptually salient than L-M or achromatic textures, so for condition b) I used the strongest S-cone texture possible (+/- 0.7 relative to the neutral colour).

Results

In experiment 2.2.1 suggested that texture-texture contrast might be able to account for the unexpected result, when I measured the strength of simultaneous contrast induction when both target and surround were textured. The results of this experiment throw this interpretation into doubt.

Figure 43 shows the results for all four experiments, for three observers. We find that even very weak L-M textures are sufficient to almost completely inhibit simultaneous contrast effects induced by a background shift along the isoluminant +L-M axis. This suggests that in the natural world, induction effects will not affect the perceived colours of objects that are even lightly textured. This may explain why simultaneous contrast effects only rarely enter into our everyday experience, even though they may be very strong in laboratory demonstrations.

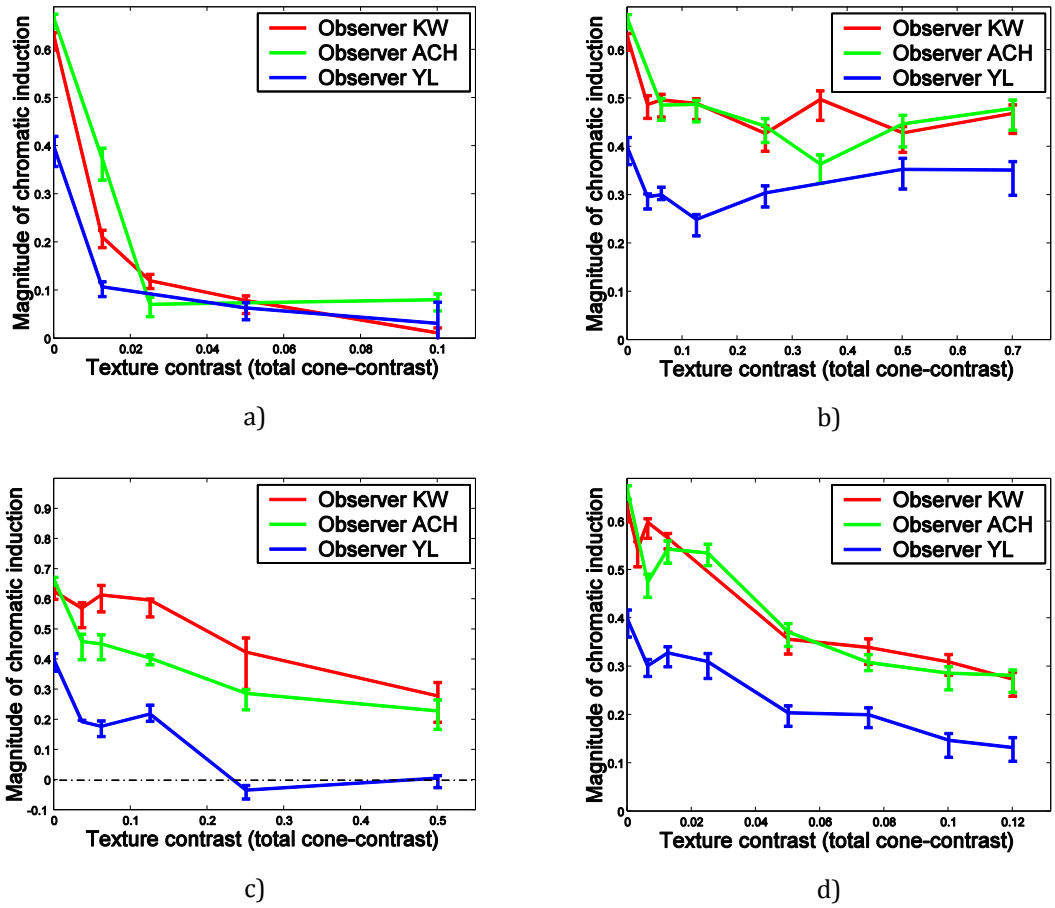


Figure 43: The effect of varying the internal texture contrast, on chromatic induction in textured stimuli. Texture-contrasts of '0' correspond to the 'uniform-surface' conditions of the other experiments.

a) Target texture varying along the L-M direction only

b) Target texture varying along the S-cone axis only

c) Target texture varying along the achromatic axis

d) Background texture varying along the L-M axis

S-cone isolating textures modulate simultaneous contrast in the +L-M axis surprisingly little, suggesting that simultaneous contrast effects may be tightly tuned to the cardinal axes. Achromatic textures influence induction somewhat more strongly than S-cone isolating textures, but their effect is still considerably weaker than that of +L-M textures. This implies that the textures must exert their influence at an opponent level, as an achromatic texture would be highly visible to a mechanism that looked at the +L or +M channels in isolation but would be far less visible to a mechanism that looked at the opponent signal.

Condition a) – L-M axis textured targets

Observer: ACH

Contrast level: 0.00000 Mean: 0.666 95% Confidence-interval 0.648 to 0.687 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.01250 Mean: 0.362 95% Confidence-interval 0.300 to 0.390 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.02500 Mean: 0.069 95% Confidence-interval 0.027 to 0.090 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.10000 Mean: 0.081 95% Confidence-interval 0.058 to 0.103 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

p-value between uniform and lowest contrast squares: 0.000

p-value between uniform and highest contrast squares: 0.020

Observer: KW

Contrast level: 0.00000 Mean: 0.626 95% Confidence-interval 0.604 to 0.642 Mean number of trials per point: 53.13 (Max = 55; Min = 51)

Contrast level: 0.01250 Mean: 0.211 95% Confidence-interval 0.189 to 0.231 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.02500 Mean: 0.120 95% Confidence-interval 0.104 to 0.145 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.05000 Mean: 0.089 95% Confidence-interval 0.068 to 0.110 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

Contrast level: 0.10000 Mean: 0.012 95% Confidence-interval -0.010 to 0.034 Mean number of trials per point: 50.00 (Max = 50; Min = 50)

p-value between uniform and lowest contrast squares: 0.000

p-value between uniform and highest contrast squares: 0.000

Observer: YL

Contrast level: 0.00000 Mean: 0.397 95% Confidence-interval 0.360 to 0.438 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

```
Contrast level: 0.01250 Mean: 0.108   95% Confidence-interval 0.088 to 0.126       Mean number of trials per point: 50.00 (Max = 50; Min =
50)
Contrast level: 0.05000 Mean: 0.064   95% Confidence-interval 0.038 to 0.087       Mean number of trials per point: 50.00 (Max = 50; Min =
50)
Contrast level: 0.10000 Mean: 0.007   95% Confidence-interval -0.390 to 0.386      Mean number of trials per point: 20.00 (Max = 20; Min =
20)
p-value between uniform and lowest contrast squares: 0.000
p-value between uniform and highest contrast squares: 0.000
```

Condition b) – S-axis textured targets

Observer: ACH

Contrast level: 0.00000 Mean: 0.666 95% Confidence-interval 0.645 to 0.685 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.06125 Mean: 0.482 95% Confidence-interval 0.450 to 0.509 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.12500 Mean: 0.492 95% Confidence-interval 0.456 to 0.508 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.25000 Mean: 0.446 95% Confidence-interval 0.411 to 0.480 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.35000 Mean: 0.369 95% Confidence-interval 0.331 to 0.410 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.50000 Mean: 0.442 95% Confidence-interval 0.394 to 0.479 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.70000 Mean: 0.480 95% Confidence-interval 0.441 to 0.516 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
p-value between uniform and lowest contrast squares: 0.000
p-value between uniform and highest contrast squares: 0.000

Observer: KW

Contrast level: 0.00000 Mean: 0.626 95% Confidence-interval 0.604 to 0.642 Mean number of trials per point: 53.13 (Max = 55; Min = 51)
Contrast level: 0.03612 Mean: 0.497 95% Confidence-interval 0.461 to 0.527 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.06125 Mean: 0.500 95% Confidence-interval 0.469 to 0.527 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.12500 Mean: 0.491 95% Confidence-interval 0.460 to 0.513 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.25000 Mean: 0.428 95% Confidence-interval 0.392 to 0.467 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

```
Contrast level: 0.35000 Mean: 0.508   95% Confidence-interval 0.467 to 0.546       Mean number of trials per point: 20.00 (Max = 20; Min =
20)
Contrast level: 0.50000 Mean: 0.430   95% Confidence-interval 0.390 to 0.463       Mean number of trials per point: 20.00 (Max = 20; Min =
20)
Contrast level: 0.70000 Mean: 0.468   95% Confidence-interval 0.424 to 0.504       Mean number of trials per point: 20.00 (Max = 20; Min =
20)
p-value between uniform and lowest contrast squares: 0.000
p-value between uniform and highest contrast squares: 0.000
```

Observer: YL

Contrast level: 0.00000 Mean: 0.397 95% Confidence-interval 0.363 to 0.438 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.03612 Mean: 0.295 95% Confidence-interval NaN to 0.303 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.06125 Mean: 0.302 95% Confidence-interval 0.286 to 0.325 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.12500 Mean: 0.249 95% Confidence-interval 0.202 to 0.269 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.25000 Mean: 0.304 95% Confidence-interval 0.270 to 0.328 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.50000 Mean: 0.355 95% Confidence-interval 0.309 to 0.398 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.70000 Mean: 0.349 95% Confidence-interval 0.305 to 0.388 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

p-value between uniform and lowest contrast squares: 0.000

p-value between uniform and highest contrast squares: 0.306

Condition c) – Achromatically textured targets

Observer: ACH

Contrast level: 0.00000 Mean: 0.666 95% Confidence-interval 0.649 to 0.686 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.03612 Mean: 0.468 95% Confidence-interval 0.400 to 0.508 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.06125 Mean: 0.459 95% Confidence-interval 0.401 to 0.506 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.12500 Mean: 0.403 95% Confidence-interval 0.378 to 0.409 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.25000 Mean: 0.295 95% Confidence-interval 0.225 to 0.334 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.50000 Mean: 0.234 95% Confidence-interval 0.181 to 0.294 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

p-value between uniform and lowest contrast squares: 0.000

p-value between uniform and highest contrast squares: 0.000

Observer: KW

Contrast level: 0.00000 Mean: 0.626 95% Confidence-interval 0.602 to 0.642 Mean number of trials per point: 53.13 (Max = 55; Min = 51)

Contrast level: 0.03612 Mean: 0.577 95% Confidence-interval 0.500 to 0.600 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.06125 Mean: 0.621 95% Confidence-interval 0.561 to 0.678 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.12500 Mean: 0.598 95% Confidence-interval 0.555 to 0.629 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.25000 Mean: 0.420 95% Confidence-interval 0.300 to 0.490 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.50000 Mean: 0.278 95% Confidence-interval 0.181 to 0.378 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

```
p-value between uniform and lowest contrast squares: 0.032
p-value between uniform and highest contrast squares: 0.000
```

Observer: YL

Contrast level: 0.00000 Mean: 0.397 95% Confidence-interval 0.360 to 0.438 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.03612 Mean: 0.199 95% Confidence-interval 0.190 to 0.202 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.06125 Mean: 0.179 95% Confidence-interval 0.139 to 0.207 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.12500 Mean: 0.222 95% Confidence-interval 0.187 to 0.265 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.25000 Mean: -0.033 95% Confidence-interval -0.068 to -0.005 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.50000 Mean: 0.005 95% Confidence-interval -0.026 to 0.021 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

p-value between uniform and lowest contrast squares: 0.000

p-value between uniform and highest contrast squares: 0.000

Condition d) – Mondrian textured background

Observer: ACH

Contrast level: 0.00000 Mean: 0.666 95% Confidence-interval 0.647 to 0.685 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.00625 Mean: 0.481 95% Confidence-interval 0.450 to 0.509 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.01250 Mean: 0.545 95% Confidence-interval 0.503 to 0.563 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.02500 Mean: 0.538 95% Confidence-interval 0.506 to 0.562 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.05000 Mean: 0.373 95% Confidence-interval 0.340 to 0.398 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.07500 Mean: 0.310 95% Confidence-interval 0.294 to 0.341 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.10000 Mean: 0.292 95% Confidence-interval 0.258 to 0.317 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.12000 Mean: 0.282 95% Confidence-interval 0.248 to 0.299 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
p-value between uniform and lowest contrast squares: 0.000
p-value between uniform and highest contrast squares: 0.000

Observer: KW

Contrast level: 0.00000 Mean: 0.626 95% Confidence-interval 0.603 to 0.643 Mean number of trials per point: 53.13 (Max = 55; Min = 51)
Contrast level: 0.00313 Mean: 0.545 95% Confidence-interval 0.513 to 0.577 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.00625 Mean: 0.600 95% Confidence-interval 0.571 to 0.619 Mean number of trials per point: 20.00 (Max = 20; Min = 20)
Contrast level: 0.01250 Mean: 0.567 95% Confidence-interval 0.546 to 0.587 Mean number of trials per point: 40.00 (Max = 40; Min = 40)

```
Contrast level: 0.05000 Mean: 0.355   95% Confidence-interval 0.320 to 0.381       Mean number of trials per point: 20.00 (Max = 20; Min =
20)
Contrast level: 0.07500 Mean: 0.333   95% Confidence-interval 0.295 to 0.365       Mean number of trials per point: 20.00 (Max = 20; Min =
20)
Contrast level: 0.10000 Mean: 0.313   95% Confidence-interval 0.287 to 0.341       Mean number of trials per point: 20.00 (Max = 20; Min =
20)
Contrast level: 0.12000 Mean: 0.272   95% Confidence-interval 0.237 to 0.296       Mean number of trials per point: 20.00 (Max = 20; Min =
20)
p-value between uniform and lowest contrast squares: 0.000
p-value between uniform and highest contrast squares: 0.000
```

Observer: YL

Contrast level: 0.00000 Mean: 0.397 95% Confidence-interval 0.357 to 0.436 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.00625 Mean: 0.301 95% Confidence-interval 0.266 to 0.319 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.01250 Mean: 0.328 95% Confidence-interval 0.302 to 0.350 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.02500 Mean: 0.311 95% Confidence-interval 0.278 to 0.346 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.05000 Mean: 0.206 95% Confidence-interval 0.180 to 0.230 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.07500 Mean: 0.204 95% Confidence-interval 0.171 to 0.225 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.10000 Mean: 0.148 95% Confidence-interval 0.110 to 0.173 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

Contrast level: 0.12000 Mean: 0.132 95% Confidence-interval 0.100 to 0.151 Mean number of trials per point: 20.00 (Max = 20; Min = 20)

p-value between uniform and lowest contrast squares: 0.012

p-value between uniform and highest contrast squares: 0.000

Table 12: the raw data used to draw figure 43. p-values are also provided for two pairwise comparisons for each observer for each condition, between the uniform target and the lowest contrast target, and between the uniform target and the most contrasty target.

Experiment 2.2.4

Aims

To determine the locus of the texture-effect in simultaneous contrast induction. In this experiment, I investigate whether interactions between textures and simultaneous induction are tightly tuned to the three cardinal axes of colour-space. Cells at low levels of the visual system are tightly tuned to the cardinal axes of colour-space; however individual cells at higher levels of the nervous system are tuned to arbitrary directions in colour-space. The results of these experiments may therefore aid in determining whether the mechanisms involved in simultaneous contrast induction are sited at an early or a high level of the visual system.

Methods

In experiment 2.2.3, we examined the relationship between the internal contrast of the target in a textured surface and the strength of the simultaneous contrast effect. Here we employ identical experimental methods but test three different backgrounds oriented along the L-M axis, S-axis, and an intermediate axis. We measure the strength of induction for five textured patches, with the texture in each aligned along a single chromatic axis including both the two cardinal chromatic axes, and three intermediate axes. If our results appear similar for background shifts along all three axes, this will tend to favour a cortical site as the locus of the texture-inhibition effect, but if the intermediate axis differs from the cardinal axes then this will suggest that both induction and the texture-inhibition effect are located at an earlier level of the visual system.

Stimuli

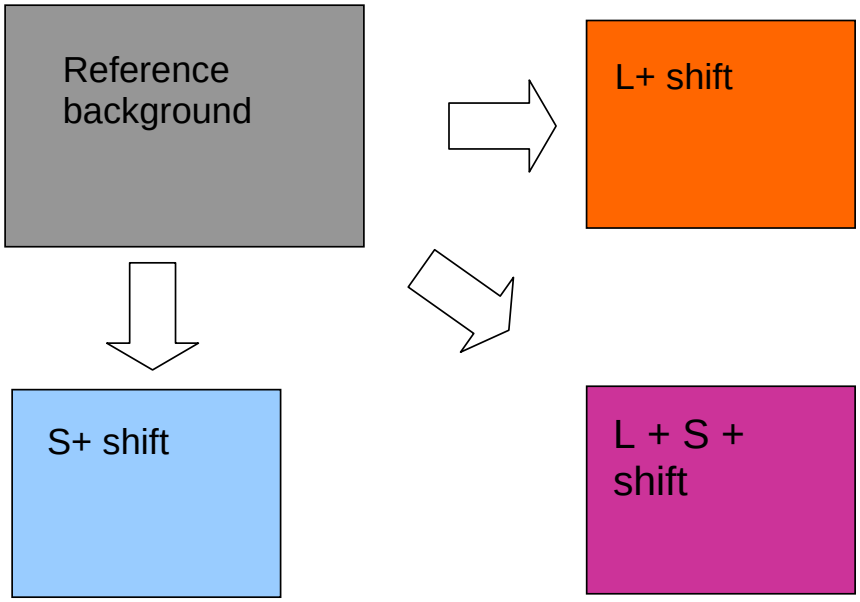


Figure 44: The three chromatic shifts used in experiment 2.2.3

Clockwise from top-right:

1) *Isoluminant +L-M shift (redder)*
2) *Isoluminant chromatic shift in a +L+S-M direction intermediate to the two cardinal axes.*
3) *Isoluminant +S shift (bluer)*

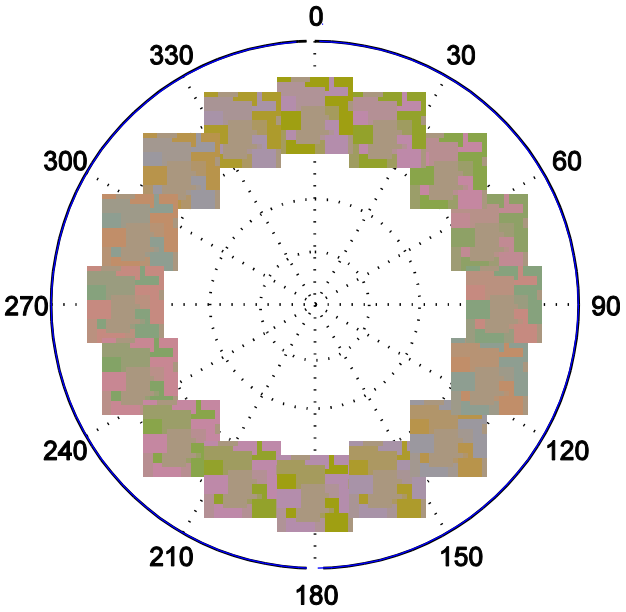


Figure 45: Textured target squares, plotted along a polar axis. The L-M axis runs horizontally through 90-270° and the (L+M)-S-axis runs vertically from 0-180°

Figure 44 and figure 45 show examples of the backgrounds and targets used in this experiment. Note that half of the targets are mirror-images of each other – e.g. the colours in the 0^{o} and 180^{o} targets are identical, but their positions in each case are complimentary. This equivalence halves the number of samples that must be taken in each case.

Stimulus dimensions and timings were identical to those specified for experiment 2.1.1, save that no mask was used for these experiments.

Equating the (L-M) and (L+M)-S axes for perceptual salience

Before starting the experiment, each observer carried out a range of observations to determine the L-M cone contrast that is equivalent to the maximum S-cone contrast available for each observer (S-cone contrast were less salient than L-cone contrasts, and observers were asked to . Observers carried this out in two ways – firstly they were asked to perform a "more" / "less" contrasty 2AFC task, where they were asked to rate the contrast of L-M targets relative to those of a fixed reference S-cone contrast – a cone contrast of 0.7 relative to neutral – this figure was chosen as it was the strongest isoluminant S-cone contrast that could be generated on my computer's monitor. In a separate task, we determined the ratio between the minimum detectable S-cone contrast and the minimum detectable L-M cone contrast, and scaled them accordingly so that the S-cone contrast level was 0.7. The values for the L-M cone texture contrast were similar for both methods, so we chose to use the results from the first method only to generate target textures that were personalised for each observer.

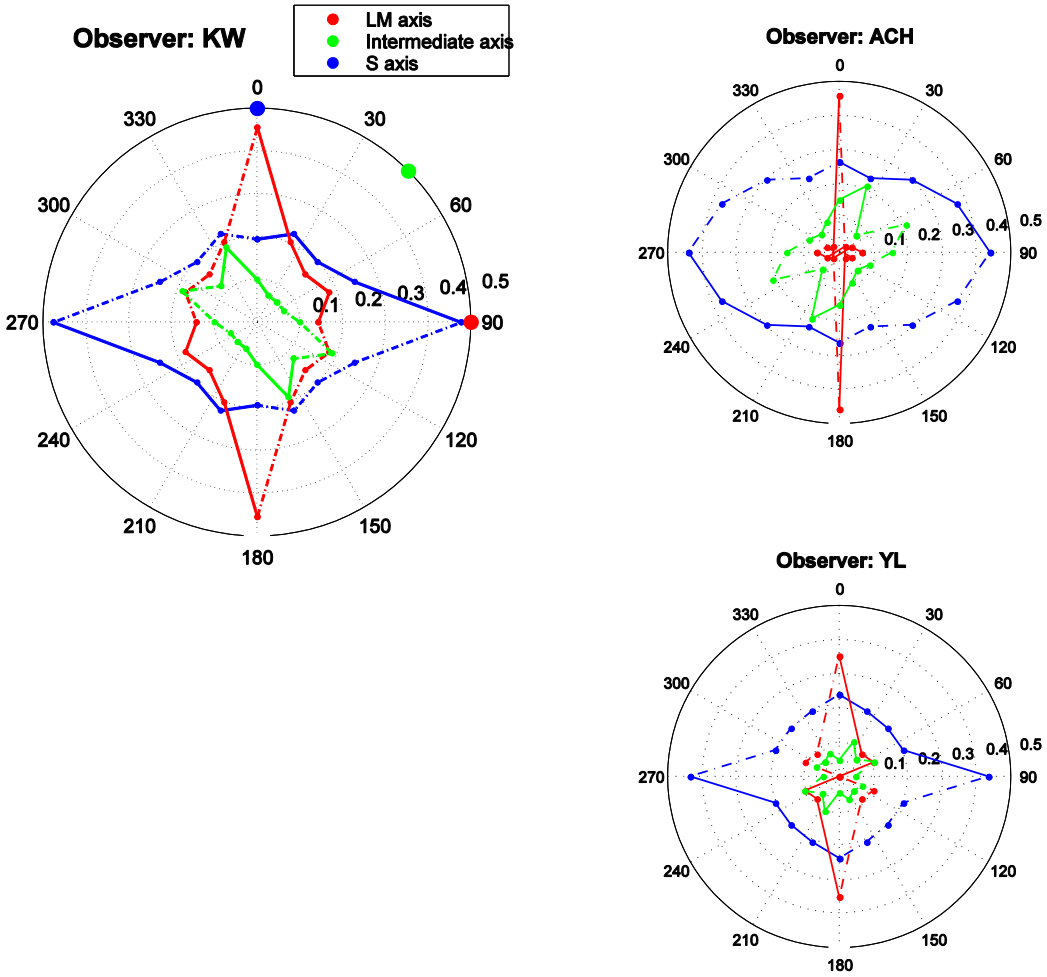


Figure 46: Polar plots for three observers show that the effect of textured stimuli on simultaneous contrast induction is closely tuned to the cardinal axes. On each polar plot, the strength of induction is indicated by the radial distance from the origin, and the chromatic axis of the texture is represented by the rotation from the origin. Note that inverse textures (e.g. red-green v. green-red) are equivalent, so many textures are duplicated and need not be investigated. The 'S' axis is vertical and the 'L-M' axis is horizontal. The direction of an intermediate axis has been marked by a green point.

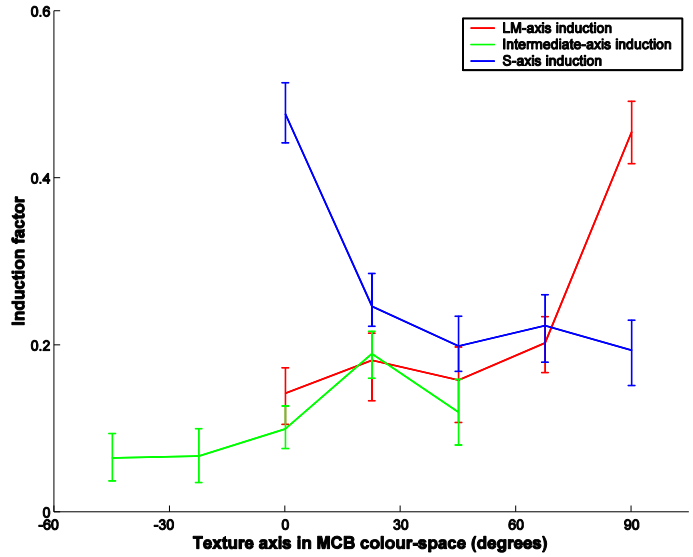
Three closed polygons are plotted for each of the three background shifts investigated: one was a pure +L shif (red)t, the next was a pure +S shift (blue) and the third was an intermediate chromatic shift (green). The strength of induction was measured for five target textures tested against each background. These values are plotted as points joined by solid lines; the dashed lines mirror the solid lines, but indicate points that were not measured.

Figure 46 shows that an S-cone modulated texture has little effect on contrast elicited by an LM-axis background shift, and vice versa. More interestingly, if the direction of the background shift is intermediate to the L and M axes then induction is always affected by either L-M texture in the target, or S-cone texture in the target, and the target texture cannot be varied along any chromatic axis that allows simultaneous contrast to act.

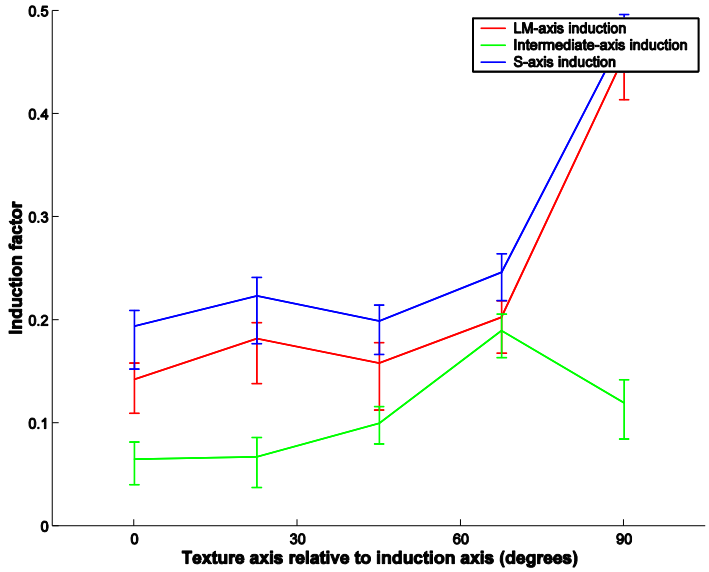
It is difficult to analyse these data statistically, for reasons outlined in Appendix F. However, it is possible to perform significance tests along the long and short axes of each ellipse, and if the resulting values were not significant then this would cast real doubt on any conclusions that are drawn from them.

Table 13 shows the results of this approach. Although it is fraught by problems such as multiple comparisons that reduce the power of the test to a low level, the results appear clear: comparisons 1-3 show that the intermediate axis is more circular than the cardinal axes, which are statistically significantly elliptical. A second conclusion that can be drawn from comparisons 4 and 5 - is that reduction is generally reduced in the intermediate axis relative to induction in the cardinal axes.

It is also possible to plot the data in a way that makes it more straightforward to provide error-bars; something that would make the current polar plots too crowded to be readily visually interpreted. Figure 47a shows how this may be performed for a single quadrant of each axis. Figure 47b shows how the graph may further be simplified, clearly showing the differences between the cardinal and intermediate axes. For the cardinal axes, there is a clear drop in induction whenever the texture contains a chromatic component along the axis of induction. Based on a comparison of the size of the overall magnitude of the effect relative to the size of the error bars, it is reasonable to argue that the pattern shown by the data is robust. If formal hypothesis testing is performed, then p-values of zero support this viewpoint, although for reasons given in Appendix F, they cannot show whether the interpretation of the data is correct.

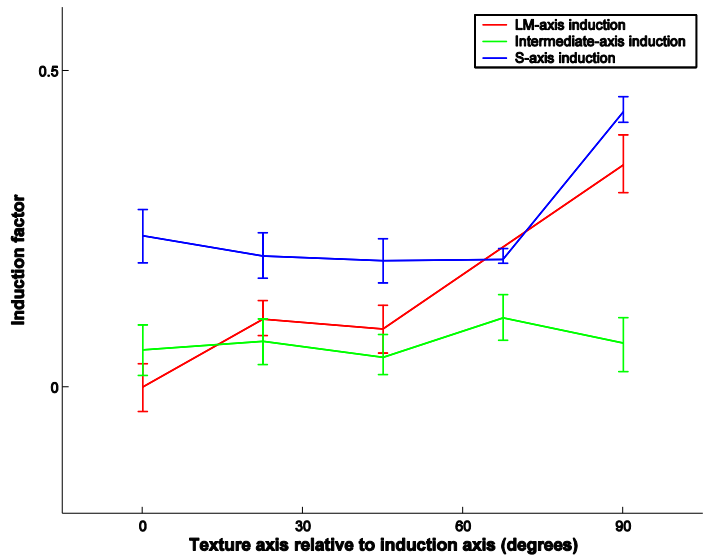


a)

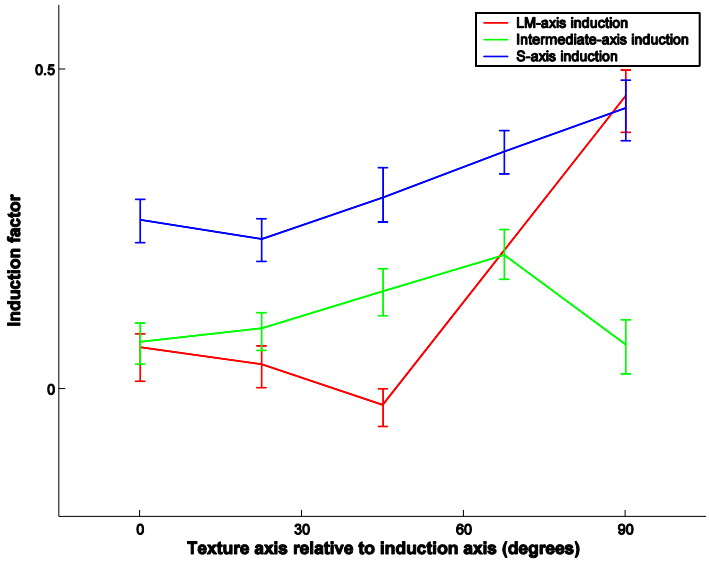


b)

Figure 47: Graphs of KW's data plotted in terms of the induction factor relative to the chromatic axis of the texture, for induction along the LM axis, S-axis and an intermediate axis. Error bars indicate 95% confidence intervals obtained using psignifit. Graph a) shows the data in an untransformed mode. Graph b) shows the data, translated and mirrored (where applicable) in such a way that all three axes can be overlaid such that they should show exactly the same pattern of data. As is clearly seen, when this is done, both the LM and S-axes appear to show a very similar pattern of data, whereas the pattern of the intermediate axis is very different: induction is weakened whatever the axis of the intermediate texture.



a)– Normalised graph for observer YL



b) – Normalised graph for observer ACH

Figure 48: normalised figures for observers YL (figure a) and ACH (figure b). Although these graphs are less clear than for observer KW, it can still be strongly argued that they illustrate the same overall pattern of data. Error-bars indicate 95% confidence intervals.

Observer KW:

Comparison 1: L-induction condition; red-green texture v. blue-yellow texture: p-value = 0

Comparison 2: S-induction condition; red-green texture v. blue-yellow texture: p-value = 0

Comparison 3: intermediate-axis induction condition; same-axis texture v. orthogonal-axis texture: p-value = 0.40822

Comparison 4: L-induction condition; blue-yellow texture v. intermediate axis; orthogonal texture: p-value = 0

Comparison 5: S-induction condition - red-green texture v. intermediate axis induction - orthogonal texture: p-value **= 0**

Observer YL:

Comparison 1: L-induction condition; red-green texture v. blue-yellow texture: p-value = 0

Comparison 2: S-induction condition; red-green texture v. blue-yellow texture: p-value = 0

Comparison 3: intermediate-axis induction condition; same-axis texture v. orthogonal-axis texture: p-value = 0.99027

Comparison 4: L-induction condition; blue-yellow texture v. intermediate axis; orthogonal texture: p-value = 0

Comparison 5: S-induction condition - red-green texture v. intermediate axis induction - orthogonal texture: p-value **= 0**

Observer ACH:

Comparison 1: L-induction condition; red-green texture v. blue-yellow texture: p-value = 0

Comparison 2: S-induction condition; red-green texture v. blue-yellow texture: p-value = 0

Comparison 3: intermediate-axis induction condition; same-axis texture v. orthogonal-axis texture: p-value = 0.76021

Comparison 4: L-induction condition; blue-yellow texture v. intermediate axis; orthogonal texture: p-value = 0

Comparison 5: S-induction condition - red-green texture v. intermediate axis induction - orthogonal texture: p-value **= 0**

Table 13: NHST tests of selected pairwise comparisons for the three observers reported in experiment 2.2.4. All observers show highly statistically significant differences for all comparisons, for all axes other than the intermediate axis. The analysis was repeated with a fixed lambda (lapse rate) value of 0.02, but this did not significantly influence the results and the data are not reported here. Tests were performed using 'pfcmp' and test a non-directional hypothesis

Observer: KW

Induction axis: L-M induction

Texture axis: 0.0 degrees
50% Threshold estimate: 0.143
95% confidence intervals: 0.106 0.172

Texture axis: 22.5 degrees
50% Threshold estimate: 0.182
95% confidence intervals: 0.135 0.218

Texture axis: 45.0 degrees
50% Threshold estimate: 0.158
95% confidence intervals: 0.111 0.199

Texture axis: 67.5 degrees
50% Threshold estimate: 0.203
95% confidence intervals: 0.167 0.236

Texture axis: 90.0 degrees
50% Threshold estimate: 0.455
95% confidence intervals: 0.418 0.495

Induction axis: (L+M)-S axis induction

Texture axis: 0.0 degrees
50% Threshold estimate: 0.486
95% confidence intervals: 0.449 0.521

Texture axis: 22.5 degrees
50% Threshold estimate: 0.247
95% confidence intervals: 0.214 0.276

Texture axis: 45.0 degrees
50% Threshold estimate: 0.200
95% confidence intervals: 0.165 0.229

Texture axis: 67.5 degrees
50% Threshold estimate: 0.224
95% confidence intervals: 0.175 0.260

Texture axis: 90.0 degrees
50% Threshold estimate: 0.188
95% confidence intervals: 0.142 0.220

Induction axis: Intermediate-axis induction

Texture axis: -45.0 degrees
50% Threshold estimate: 0.076
95% confidence intervals: 0.045 0.100

Texture axis: -22.5 degrees
50% Threshold estimate: 0.065
95% confidence intervals: 0.029 0.096

Texture axis: 0.0 degrees
50% Threshold estimate: 0.101
95% confidence intervals: 0.076 0.128

Texture axis: 22.5 degrees
50% Threshold estimate: 0.190
95% confidence intervals: 0.158 0.219

Texture axis: 45.0 degrees
50% Threshold estimate: 0.131
95% confidence intervals: 0.091 0.171

Observer: YL

Induction axis: L-M induction

Texture axis: 0.0 degrees
50% Threshold estimate: 0.001
95% confidence intervals: -0.038 0.035

Texture axis: 22.5 degrees
50% Threshold estimate: 0.108
95% confidence intervals: 0.082 0.139

Texture axis: 45.0 degrees
50% Threshold estimate: 0.092
95% confidence intervals: 0.053 0.129

Texture axis: 90.0 degrees
50% Threshold estimate: 0.352
95% confidence intervals: 0.309 0.399

Induction axis: (L+M)-S axis induction

Texture axis: 0.0 degrees
50% Threshold estimate: 0.432
95% confidence intervals: 0.406 0.449

Texture axis: 22.5 degrees
50% Threshold estimate: 0.203
95% confidence intervals: 0.193 0.219

Texture axis: 45.0 degrees
50% Threshold estimate: 0.206
95% confidence intervals: 0.168 0.239

Texture axis: 67.5 degrees
50% Threshold estimate: 0.222
95% confidence intervals: 0.185 0.259

Texture axis: 90.0 degrees
50% Threshold estimate: 0.239
95% confidence intervals: 0.194 0.280

Induction axis: Intermediate-axis induction

Texture axis: -45.0 degrees
50% Threshold estimate: 0.070
95% confidence intervals: 0.023 0.106

Texture axis: -22.5 degrees
50% Threshold estimate: 0.110
95% confidence intervals: 0.074 0.144

Texture axis: 0.0 degrees
50% Threshold estimate: 0.048
95% confidence intervals: 0.009 0.074

Texture axis: 22.5 degrees
50% Threshold estimate: 0.072
95% confidence intervals: 0.038 0.105

Texture axis: 45.0 degrees
50% Threshold estimate: 0.063
95% confidence intervals: 0.022 0.100

Observer: ACH

Induction axis: L-M induction

Texture axis: 0.0 degrees
50% Threshold estimate: 0.066
95% confidence intervals: 0.011 0.085

Texture axis: 22.5 degrees
50% Threshold estimate: 0.039
95% confidence intervals: 0.004 0.069

Texture axis: 45.0 degrees
50% Threshold estimate: -0.025
95% confidence intervals: -0.059 -0.000

Texture axis: 90.0 degrees
50% Threshold estimate: 0.459
95% confidence intervals: 0.403 0.500

Induction axis: (L+M)-S axis induction

Texture axis: 0.0 degrees
50% Threshold estimate: 0.440
95% confidence intervals: 0.391 0.485

Texture axis: 22.5 degrees
50% Threshold estimate: 0.372
95% confidence intervals: 0.337 0.406

Texture axis: 45.0 degrees
50% Threshold estimate: 0.300
95% confidence intervals: 0.259 0.337

Texture axis: 67.5 degrees
50% Threshold estimate: 0.235
95% confidence intervals: 0.200 0.268

Texture axis: 90.0 degrees
50% Threshold estimate: 0.265
95% confidence intervals: 0.230 0.299

Induction axis: Intermediate-axis induction

Texture axis: -45.0 degrees
50% Threshold estimate: 0.070
95% confidence intervals: 0.016 0.104

Texture axis: -22.5 degrees
50% Threshold estimate: 0.210
95% confidence intervals: 0.175 0.247

Texture axis: 0.0 degrees
50% Threshold estimate: 0.154
95% confidence intervals: 0.116 0.190

Texture axis: 22.5 degrees
50% Threshold estimate: 0.096
95% confidence intervals: 0.064 0.122

Texture axis: 45.0 degrees
50% Threshold estimate: 0.074
95% confidence intervals: 0.037 0.103

Table 14: raw data and confidence intervals for experiment 2.2.4

This evidence supports the hypothesis that both simultaneous induction, and the inhibitory effects of texture contrast occur entirely at a low level of the visual system, as at levels higher than V1, few neurons can be found that respond exclusively along a single axis of the visual system as they do at lower levels such as the LGN or retina. If simultaneous contrast induction were processed at a high level of the visual cortex where the cells are not tuned along the cardinal axes, it would be difficult to explain the strong tuning exhibited here. Similarly, if texture were processed in a higher cortical area then we would be forced to explain how and why untuned texture processing mechanisms are able to feed back to highly tuned mechanisms, presumptively sited at a lower level of the visual system. For an in-depth review of the relevant anatomical and physiological evidence, please see chapter 2 – the organization and function of the visual system.

Discussion and conclusions

The colour-appearance of multi-coloured objects has rarely, if ever been studied in the laboratory. Here we show that simultaneous induction works differently in textured objects and in the uniform objects that have hitherto been investigated. This manner is consistent with our hypothesis that simultaneous contrast acts more strongly between attached than between detachable surfaces. Furthermore, these results shed some light on the question posed in the introduction: why is it that simultaneous contrast effects are not obvious in real-world situations? The practical significance of my findings here is that because most real-world objects display surface texture, they will not undergo significant simultaneous chromatic contrast with their surroundings. Whether the inhibition of simultaneous contrast effects is a high-level effect which requires visuo-cognitive mechanisms underlying scene segmentation, or a low-level effect, which might occur in the retina via inhibition of at edges, is a question that remains unanswered.

Chapter 6, Experiment 3 – Motion and Binocular Disparity

Introduction

In addition to texture differences, relative motion and relative depth between figure and background may both serve as powerful cues to detachability of the figure from the background. The experiments reported in this chapter aimed to investigate whether these two cues influenced the strength of simultaneous contrast induction. They are also examined together because of the similarities in the methodology used to measure induction in each experiment. Rather than employing a 2AFC technique as for the previous experiments, here we employ nulling methods.

The main hypothesis to be tested is that detachability should reduce the influence of an inducing background. Both motion and relative depth provide strong cues to detachability, and if my working hypothesis that detachability should reduce the strength of simultaneous induction is correct then the introduction of both cues should act to reduce the test surface's dependence on the surround. The second main objective of these experiments is to indicate the likely locus of the simultaneous contrast mechanisms.

Motion

Motion is a strong cue that can segment figure from ground (Regan and Beverley 1984; Moller and Hurlbert 1996), and to the best of my knowledge, no previous experimenters have investigated whether motion affects simultaneous contrast induction. As discussed in chapter 2, there are at least two forms of motion perception – first order and second order. It seems likely that first order motion is represented early in the visual system – perhaps even in the retina. However, second order motion processing is thought to involve extrastriate areas. As determining the likely location of simultaneous-contrast mechanisms is one of my prime objectives, it is desirable to investigate both first and second order motion.

Regan and Beverley have demonstrated how motion can cause previously 'camouflaged' objects to suddenly become visible (Regan and Beverley 1984). One objection to the concept that motion contrast should influence colour perception, could be to point out that this should cause a previously unnoticed object to appear to suddenly change colour. However, similar objections could be made for other effects, such as shading on geometrically bistable surfaces. Inattentional blindness could account for our general failure to notice such changes (Simons and Chabris 1999).

Depth

Bela Julesz's stereograms showed shapes that were defined only through stereoscopic disparity. As such, binocular disparity is clearly a strong cue to depth perception. As discussed in Chapter 1 (Introduction), many experiments have investigated the relationship between simultaneous induction and perceived depth, though their results were mixed. Some experimenters found little or no effect, but others obtained very significant results. Although the vast majority of these experiments were performed in the achromatic domain, Shevell and Miller (Shevell and Miller 1996) recently investigated whether perceived depth could influence colour perception (finding a weak but consistent effect), and Kinjiro Amano (Amano, Foster et al. 2002) and I presented similar posters at the same ECVP session, investigating links between simultaneous contrast induction and depth perception using a 2AFC technique and a nulling technique respectively.

Why have researchers obtained such differing results, when they investigated links between simultaneous contrast induction and perceived depth? One reason might have been in the number and nature of depth cues available in different investigations. A multitude of depth cues are available, although we are generally capable of perceiving depth even when only a few of the cues are present.

A brief survey of the literature shows that differences in the type and number of depth cues probably do not resolve the issue: two of the papers to have found an effect were that of Shevell and Miller (Shevell and Miller 1996), and Gilchrist (Gilchrist 1977). Shevell and Miller's experiment was performed on a computer monitor and the only depth cue available to their observers was that of stereoscopic depth, which in isolation does not replicate many of the features of depth in 3d objects such as optical blurring or changes in relative size – although this could probably be interpreted as showing that any effect observed is truly due to perceived depth rather than changing physical attributes of the stimuli. Gilchrist's experiment found a very large effect of perceived depth, but his stimuli were viewed monocularly, preventing observers from relying on convergence cues, motion parallax or stereoscopic disparity in making their judgments of depth. He influenced perceived depth by varying only one cue – interposition – and focal cues such as optical blurring between different depth planes were in conflict with perceived disparity. Conversely the majority of researchers – such as Gibbs and Lawson (Gibbs and Lawson 1973) and Mershon (Mershon 1971), who respectively found no effect and a tiny effect, used physical stimuli that provided most depth cues – such as stereoscopic disparity, size cues and motion parallax. Gilchrist argued that the important difference between his experiment and preceding experiments that found little or no effect was that his stimuli provided alternative illumination frameworks. Although his stimuli contained depth cues, they signaled which of the available illumination frameworks the tests surface should belong to, rather than simply indicating whether two surfaces 'belonged'. Although it would be interesting to recreate Gilchrist's experiments using coloured lights, this is beyond the scope of my current line of investigation.

Aims and predictions

Prior to commencing the series of experiments, I did not make any predictions regarding the relative likelihoods of the two main outcomes. In Platt's terminology (Platt 1964), I envisaged the series of experiments essentially as 'critical tests' of the notion that detachability influences contrast induction. I remained genuinely agnostic as to the probable outcome, an approach regarded as favourable by many as it prevents scientists becoming too attached to 'pet' theories, or failing to consider alternatives (Chamberlin 1965).

In advanced sciences, such as physics, it is often possible to make extremely precise, quantitative predictions. Unfortunately, the biological sciences are generally 'messy' to a degree that precludes this, even when there are relatively sophisticated models of the biological processes in question. As I have not presented any complex models of simultaneous contrast induction, any predictions I make are therefore simple inequalities that can generally be outlined very briefly.

Methods

Rationale for using a nulling method

For various reasons it would be difficult to investigate either motion or depth cues using the two-alternative forced choice method used in experiments 1 and 2.

In pilot experiments I found that it was difficult to perceive the stereoscopic disparity cues accurately if the stimuli were presented for short durations. This could be partly overcome by making the stimuli last longer. However, trial-based experiments become inefficient if the trials last too long – and I had residual concerns that stereo perception may be disrupted by the abrupt background colour change that is fundamental to the 2AFC method.

In motion experiments, the problem is more serious again. Motion is a change in spatial position over time, yet the 2AFC method aims to determine the strength of simultaneous contrast as induced by an instantaneous background change.

For these reasons, I felt I needed to use a different method in order to investigate interactions between simultaneous contrast induction and differential motion or depth. However I still wished to avoid eliciting long-term adaptation or using a paradigm that involves medium-term colour memory, which may be implicated when certain methods such as dichoptic matching are used.

The nulling technique

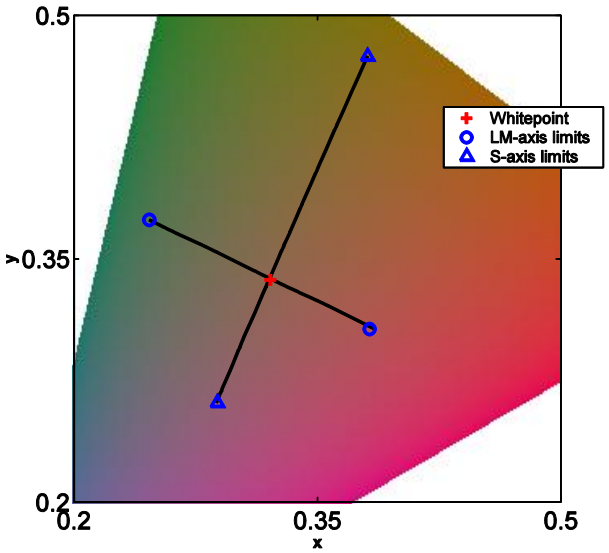


Figure 49: This CIE chromaticity diagram shows the L-M and (L+M)-S chromatic axes used for the motion and disparity experiments discussed in this chapter, plotted against a triangle delineating the gamut of the monitor. Both axes share the same neutral 'white-point' whose CIE x and y co-ordinates were 0.321 and 0.337 respectively.

In this set of experiments, we use a different method of quantifying simultaneous contrast induction from the 2AFC technique used for previous experiments in this thesis. The technique is based on a nulling paradigm first used by Krauskopf, Zaidi and Mandler in 1986 (Krauskopf, Zaidi et al. 1986) and later used by Zaidi et al. (Zaidi, Yoshimi et al. 1991; Zaidi, Yoshimi et al. 1992; Zaidi and Zipser 1993) where simultaneous contrast is induced by modulating the background colour sinusoidally along an arbitrary chromatic axis (figure 49), as shown in figure 50. This method typically induces very strong changes in colour appearance in the test objects, and because the background colour modulation occurs at a relatively rapid rate, long-term chromatic adaptation is minimised. However, this method shares a drawback of the previous experiments: namely that the background changes may induce temporal contrast adaptation, whose time course is presently unknown.

In this nulling paradigm, the modulation of the background colour proceeds at a fixed rate and magnitude throughout the experiment. At any given instant during the experiment, the colour of the target square is related to the colour of the inducing background by the formula:

Target_contrast = k x Contrast_of_the_background

where 'k' is an observer-controlled variable that I call the 'induction factor' and 'contrast' refers to the total-cone-contrast between either the target or background colour and a neutral reference. The observer's task is to adjust the variable 'k' in order to stabilise the appearance of the target square. If $k = 0$, the chromaticity of the target square will remain constant, identical to the neutral reference chromaticity. Yet chromatic induction will cause the perceived colour of the target square to vary in counterphase to the colour of the background. By adjusting 'k' to a positive value dependent on the strength of induction, this induced colour change may be nulled, and the appearance of the target square will stabilise. During an experimental session, the observer is repeatedly asked to adjust 'k' until the colour of the target square stops changing. After each setting, 'k' is reset to a random value between -1 and 1, and the observer starts the next match.

An assumption made by this technique is that the relationship between the contrast of the inducing background and the magnitude of simultaneous contrast induction is linear. Although this assumption is unlikely to be absolutely true, observers reported that they find they can make quite satisfactory settings, even though it was impossible to completely stabilise the appearance of the target square. Before performing future experiments, it may be worthwhile to investigate whether relating the target colour to the background colour by a non-linear equation might allow observers to find more satisfactory matches.

As in experiments 1 and 2, we chose to use a modified McLeod-Boynton colourspace that allowed us to isolate the three cardinal chromatic and brightness axes of the early visual system. I chose moderately strong colour-contrasts that were within the gamut of the monitor, as shown on figure 49. I performed my primary experiments along the LM axis, but in control experiments I investigated induction along the other cardinal axes. However, unlike in previous experiments, I did not consider it necessary to perceptually equate the contrasts along different axes because the primary objective of my experiment is simply to determine whether or not induction occurs under particular circumstances. A comparison of the strength of simultaneous induction along different chromatic axes has little relevance to this aim.

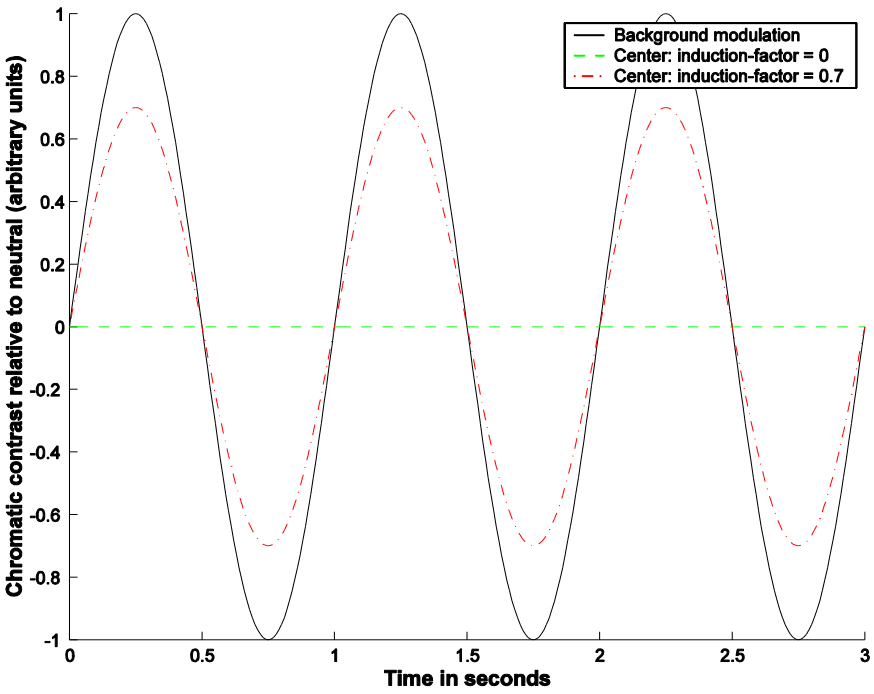


Figure 50: The chromaticities of the target and background stimulus components over time during a nulling experiment.

The x-axis marks out time from the start of the stimulus display. The y-axis indicates chromatic contrast relative to the neutral chromaticity. It is specified in arbitrary cone-contrast units. Three lines are marked on this graph:

1. *The solid black line shows the colour of the background. The magnitude of this modulation is set at a fixed level by the experimenter that is identical for all of the observers who completed each experimental condition.*
2. *If the chromaticity of the target patch were fixed – as shown by the green dashed line – simultaneous contrast induction would cause the appearance of the target to shift in the opposite direction from that of the background.*
3. *The red dashdotted line illustrates how the target chromaticity might vary in phase with the background chromaticity for a potential observer-setting. (induction factor k = 0.7). This colour-change counterbalances the induced change from the background, causing the target to retain a roughly neutral chromaticity over the course of the experiment.*

Apparatus

Computer, monitor and input device

I used a single computer for these experiments, equipped with an NVidia Geforce II graphics card with 64Mb of memory. The screen resolution was set at 1024 x 768 x 32 bits per pixel, and the vertical refresh rate was 85 Hz. All experiments were carried out on an HP 21" monitor with a Trinitron flat-screen. A simple chinrest was used to fix the observer's viewing distance at 57cm from the CRT screen.

Observers used a standard gamepad to interact with all of these experiments. The gamepad used four buttons to adjust the induction factor '*k*'. Two were used for fine control, and two for coarse adjustments. Adjustments were made by holding down buttons on the gamepad, rather than pressing them multiple times. When the observer was satisfied with a setting he pressed and released two further buttons simultaneously, at which point 'k' was randomised and the observer started the next match.

Stimulus verification

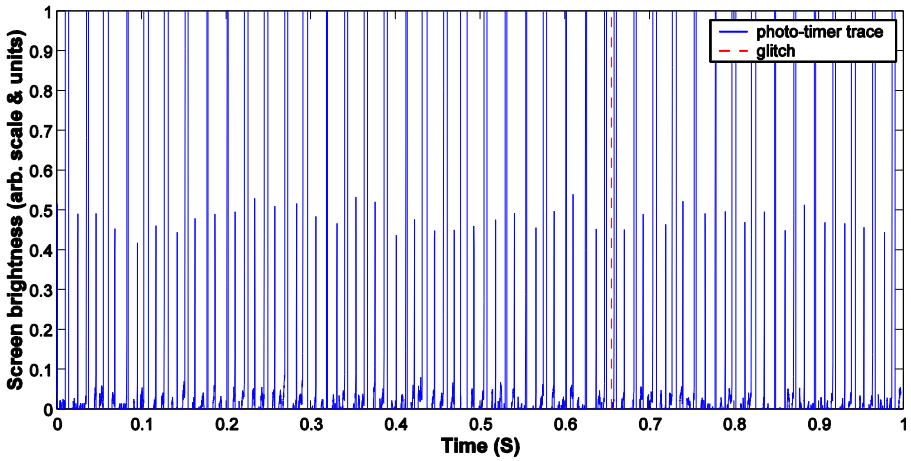


Figure 51: The output from the photodiode-based timer, used to monitor timing-glitches in the experiment.

Ideally, alternate traces should be high (bright screen) then low (dark screen). If two consecutive traces are the same height, this signifies that two consecutive scans of the screen were identical. This would mean that the moving target or background would appear to stop for a short time, causing a discontinuity in the motion.

To determine the number and severity of glitches in the experiment, the photodiode device described in Appendix G was used, in conjunction with the following short piece of code that caused the target square to be coloured either black or white on alternate frames. This small change to the stimuli took a negligible amount of time to process, and is unlikely to have significantly affected the output program.

```
#ifdef glitchtest

   if (frame_no++%2)  center_square_colour = black;

         else  center_square_colour = white;

   Set the target square position to the screen center;

#endif
```

Figure 51 shows how the data from the phototimer may be visualised in order to examine the display for bugs. Individual dropped frames are relatively common, but glitches lasting for longer than two frames are very rare.

An important note is that the experimental programs don't work as slide-shows that must show every frame in an animation. The position of the target and background, and the colour of the inducing background for each frame, are calculated individually during the experiment, based on the time elapsed from the start of the experiment. This means that glitches do not affect the frequency of the changes in the inducing backgrounds, and it also makes them less perceptible. A two frame-glitch is equivalent to a brief reduction in the frame rate to 42.5 Hz, not to a 1/85 second delay in the stimuli.

Experimental logistics

These experiments were conducted over the course of 18 months, on an ad-hoc basis. With the exception of observers KW, ACH, YZU and PMc, the observers here were recruited primarily to take part in independent experiments based around the ‘method of constant stimuli’ where it is typically necessary for the experimenter to examine preliminary data for each observer then adjust the sampling levels to match. This process, which can take five to ten minutes, left chunks of time during which observers undertook the experiments presented here.

Different amounts of time were available, depending on the number of iterations required to prepare the method-of-constants experiments. This is the main reason that observers differ in the amount of data they have each collected, and why not all observers completed all experiments. This does not matter when calculating descriptive statistics, but greatly complicates the statistical analysis of the resulting data.

Data analysis

Outliers

In early versions of the experiment if an observer pressed the 'next trial button' for too long, the experiment might rapidly accept several randomised values of *'k'* before the observer released the button. In order to discount these data, we excluded all settings more than 2 standard-deviations away from the mean. In later versions of the experiment, this bug was fixed by adding a time delay, and waiting for two buttons to be pressed and released simultaneously.

Means

The measure of contrast induction for each observer is the arithmetic mean setting of their matches for each condition.

Confidence intervals

The error bars plotted on each graph in this chapter indicate the 95% confidence intervals of the matches for each condition for each observer. They were determined by multiplying the standard errors of the mean of each dataset by 1.96.

Null-hypothesis significance-testing and p-values

One-way near-balanced ANOVA was used to provide p-values to indicate whether each observer's results for the various conditions in each experiment differed significantly from each other. This test is equivalent to a two-tailed T-test. 2-way unbalanced ANOVA was performed in order to compare conditions and observers. Although my working hypothesis is directional, and although it is possible to correct p-values obtained through ANOVA in order to evaluate directional hypotheses, in the relatively exploratory context of my own research I have not done so: for a discussion of issues surrounding directional tests see Levine and Banas (Levine and Banas 2002).

ANOVA was the obvious choice of statistical test for this data, as the raw data are inherently continuous in nature. However, a number of issues remained: should I use parametric or non-parametric ANOVA, and how should I combine data from multiple observers? The decision is further complicated by the fact that observers differ in the numbers of observations that they have made, meaning that the data are unbalanced. One possibility is to discard observations at random until each observer has the same amount of data as the observer with the smallest number of observations; however if possible it would be far preferable to retain this information and make use of unbalanced methods capable of analysing it. Finally, should I apply a correction for the reasonably large number of significance tests performed in this chapter, in order to limit experimentwise error?

A nested design of ANOVA was considered, but rejected on the suggestion of a statistician because of the unbalanced design. A type-III sum of squares model was chosen because it enables us to examine interaction effects, unlike the type-II sum of squares model, which might otherwise have superior power (Lewsey, Gardiner et al. 2001; Langsrud 2003). However, the type-III sum of squares suffers from the drawback that both main effects (in this case, observer differences and strength of induction between conditions) cannot be interpreted independently (Shaw and Mitchell-Olds 1993). In other words, if a significant difference is found between observers, as I would expect, this could also lead to a significant finding being reported between conditions even if none exists.

In many cases within this chapter, we do not find a significant between-conditions effect. However, as discussed in chapter 4, a failure to obtain statistical significance is not synonymous with proving that no difference exists. Equality testing is not performed for reasons also explained in that chapter. Neither a statistically significant nor a statistically insignificant finding can be straightforwardly interpreted for the experiments presented in this chapter. The alternative is simply to attempt to subjectively interpret the results, based on the apparent effect size, and the qualitative uniformity – or lack thereof – between observers.

Tests for Normality

To use standard ANOVA, the data should ideally satisfy a number of criteria pertaining to their distribution. I have run these tests independently for each condition measured during the following experiments, and the results for each test are included in the tables of data reported for each sub-experiment.

The most important of these, is that the data are normally distributed. There is considerable controversy both about how best to assess normality, and at least an equal amount about how to use these assessments once they have been performed. If the data is clearly not normally distributed, then it would be better to use a nonparametric method such as the Kruskallis-Wallis test; the problem is that methods used to measure normality are very sensitive, and can often indicate with a high degree of statistical significance that a dataset is not normally distributed, even when its deviation from normality is so small that it would be perfectly valid to analyse it using a test such as ANOVA that assumes normality.

The methods used to test whether data fits the normal distribution curve, are based on mathematical tests of its 'moments around the mean' – namely variance, skew and kurtosis. More advanced methods – such as the 'Z'-ratio are generally calculated using these simple variables, so I have reported them individually here for each condition for each observer, on the grounds that they can be used as the basis for more advanced calculations if desired.

Many programs normalise values of kurtosis, by subtracting 3, so that the value of kurtosis for data that are normally distributed is 'zero'. The Matlab function for kurtosis does not do this, and the values I have reported here should tend to '3' for normally distributed data.

Some cutoffs that have been suggested, are ''+/- 2' for skewness and '1 to 5' for kurtosis. For most of the data reported here, the values for both parameters are very comfortably within these boundaries. Therefore, these data appear to be normally distributed.

Homogeneity of variance

Aside from skewness and kurtosis – which tell us whether individual datasets are normally distributed – when comparing two or more conditions it is necessary to check whether they have equal variances – that is, whether they are homoscedastic. For the following datasets, I have made two sets of comparisons to match the analyses that I plan to perform: one on an individual basis for each observer, and one comparing both observers and conditions. For individual observers I report the F-statistic for homoscedasticity, as determined by Levene's test. If it is <0.05 then the requirement of homoscedasticity for that observer has not been met, at that level of significance. The F-statistic for the comparison between observers is provided below the other data, together with a short statement explaining whether the requirement for homoscedasticity has been met.

For the most part, individual observers have identical variances for each condition. However, this is not true for comparisons between observers – perhaps unsurprisingly as amongst other factors the amount of variance in an observer's settings is probably related to the amount of experience they have performing the experimental task, and may also be modified by their judgments about what constitutes an acceptable match.

The question then is: having proved that for the inter-observer comparisons the requirement of homoscedasticity is not met, should this modify the analysis that we perform? A problem with the test is that it is able to reliably detect violations of homoscedasticity at levels that are far too small to significantly impact the validity of ANOVA performed on the data. For this reason, it is desirable to examine the variances independently, and only revert to a non-parametric test if they are 'different by orders of magnitude'. Confidence intervals are stated individually for each observer for each condition, and a cursory examination suggests that for the most part, they are actually relatively close, differing at most by a factor of 2-3.

It is worth bearing in mind that violations of homoscedasticity are likely to cause a type I error (false positive) rather than a type II error (false negative). Many of the results in this chapter are negative, adding weight to the assertion that any strongly significant differences are real, and not artifacts of inter-individual differences in variance.

Cohen's 'd' statistic

Cohen's 'd' statistic is calculated and reported, both for individual observers and as a mean for each condition. The mean is weighted to accord more significance to observers with more trials. In this chapter it is possible to calculate a genuine 'd' statistic, due to the continuous nature of the data. Whenever 'd' is referred to in this chapter, the reference is to Cohen's 'd' statistic.

Standard error of the mean

The abbreviation 'SEM' is used to refer to the standard error of the mean, to save space in tables.

Standard deviation

The abbreviation 'STD' is used to refer to the standard deviation, to save space in tables.

Observer inclusion and exclusion criteria

The primary inclusion criterion for each experiment was a successful result on the Farnsworth-Munsell 100-hue test. Observers were required to wear spectacles if they normally used them for reading. Most observers were University students in their first or second year; one was a school-student (observer JH) and observers ACH and PMc were my supervisor and a member of the departmental clerical staff respectively. Observers were also required to be able to understand spoken instructions in the English language.

Data is included from all observers who attempted the experimental paradigm. Initially each observer started by running the LM-contrast (red-green) condition for a variable number of practice trials as required. Most observers found the experiment relatively straightforward but a few (PMc, AM) did not, and these observers were not asked to carry out additional experiments. The data of observer 'JH' suggested that he may be unusually strongly subject to the simultaneous contrast effect. Unfortunately, he was not available for further experiments. Training data from other observers suggests a possible explanation for his effect.

I prefer to show data from these two observers, in part because it avoids needing to justify formal post-hoc exclusion criteria that have an inherent potential to bias my results. Secondly, I believe that that the majority of the data point to a robust conclusion, despite observer JH's unusual data. However, both observers' data were removed prior to analysis by ANOVA.

Experiment 3.1: does simultaneous contrast affect moving objects?

Methods

Stimuli

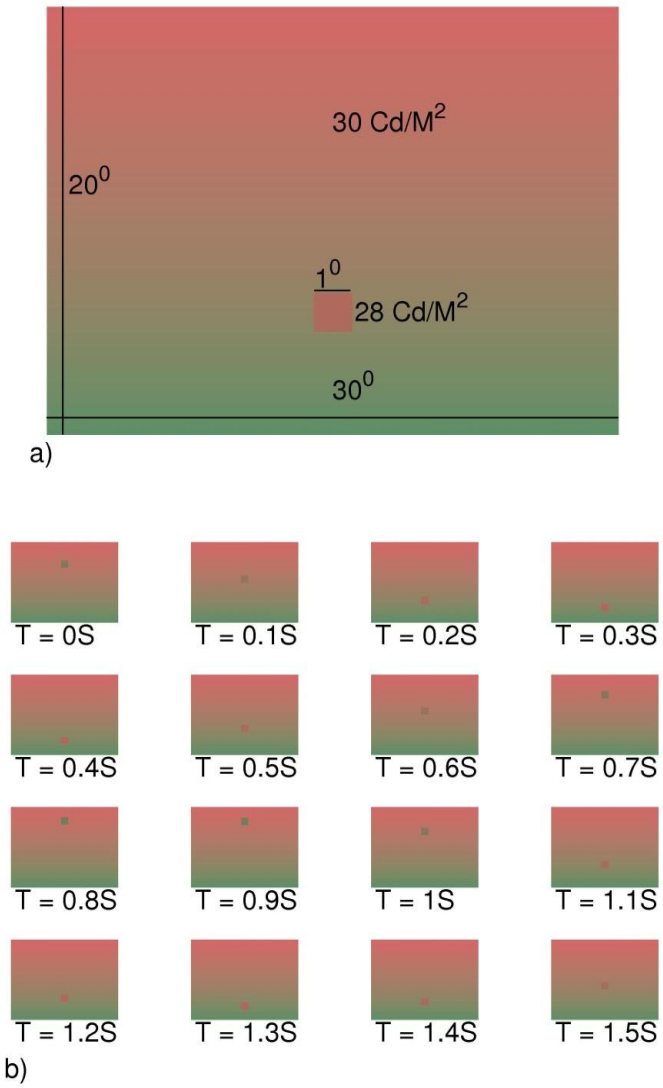


Figure 52: This figure illustrates (a) the basic stimulus parameters and (b) the motion of the target (the small square) against the fixed background.

a) The background subtends 30^o x 20^o of visual angle and has a brightness of 30Cd/M2. The brightness of the target square is set 7% dimmer to increase its contrast against the background. It subtends 1^o of visual angle.

b) Screenshots from the stimuli (1Hz condition) taken at constant intervals. The target square moves up and down along a vertical path passing through the center of the screen. Its trajectory is sinusoidal, so it spends most of the time against the saturated red or green parts of the background towards the top and bottom of the computer screen.

The center square moves along a path 2/3 the height of the screen – i.e. covering a total of 13.3 degrees of visual field, with a minimum of 3.3 degrees of visual field between the upper and lower borders of the target square and the edge of the screen.

This experiment was originally derived from a simple animation I made to demonstrate simultaneous chromatic induction, on which a small neutrally coloured square moved sinusoidally against a static chromatic gradient.

Figure 52a shows the stimuli for this experiment, which filled the entire screen of the computer monitor, and subtended 30^o x 20^o of visual angle. The background was equiluminant, with a brightness of $30Cd/M^2$. To ensure the target never vanished into the background when their mean chromaticities were identical, its brightness was set 7% dimmer than that of the background (LM and S-axis chromatic induction conditions) or redder or greener than the background (brightness induction).

The target square was always viewed against a chromatic gradient, and pilot observers were able to see a slight chromatic gradient that was induced in the target square. Observers could create a counter-gradient on the target square, to make its appearance more uniform, however the effect was subtle and few observers adjusted the stimuli to counteract it. This effect is not discussed further here.

In a sample run of the experiment, of 174.336 seconds duration and comprising 14817 frames at a frame-rate of 84.99 Hz, 214 glitches of 2-frames duration were detected, and none were detected that lasted any longer than this.

Observer directions

Observers were not instructed to fixate on a single point on the screen; indeed this would have been difficult as no fixation point was provided. Instead, they were asked to follow the target square as it moved up and down on its regular path. An unfortunate side-effect of this was that the distance from the center of fixation to the top and the bottom of the screen varied.

Predictions

Having already performed the experiment in an 'informal' manner, my expectation was that induction would be reasonably strong. Could this prediction be made more precise? The most similar experiments, in terms of the nulling paradigm used to measure induction, were those of Zaidi et al., who obtained induction-factors of between 0.6 to 0.9 for the red green axis, and 0.2-0.7 for the blue-yellow axis (Zaidi, Yoshimi et al. 1991; Zaidi, Yoshimi et al. 1992). However, Zaidi only reports data for four observers, and in any case his stimuli differed from mine in several important respects such as brightness, and the size and shape of the background and target objects. My experiment was designed to make best use of the monitor and viewing box available to me, and so no direct comparison can be made between his experiments and mine.

Another prediction could invoke data from my own experiments, reported in the previous chapters, where the induction factors for uniform background and target conditions range from about 0.5 to 0.7. Again, the stimulus conditions were not chosen to make the experiments directly comparable those in previous chapters and any attempt to do so would be futile, because of the difference between the nulling and 2AFC techniques employed in each case.

Result	*Interpretation*
0-0.2	Little or no induction present
0.2-0.5	Induction weaker than predicted. Consistent with an interpretation that the strength of induction is slightly to moderately diminished in this context.
0.5-0.7	Induction as strong as could reasonably be expected
0.7-1	Induction stronger than expected

Table 15: interpreting the results of the experiment

Table 15 shows how I would be likely to interpret different ranges of results, based on my previous findings. The cutoffs in this table are clearly partially subjective and open to debate, however I think they are reasonable in light of the previous data.

How much might one expect differential motion to reduce chromatic induction? My view is that the issue is open to debate, and no firm predictions can be made. It could be argued that differential motion is an absolute cue to detachability, and so any degree of differential motion should reduce induction to zero. However, when compound objects are considered, this is not absolutely true. A person turning their head may cause their nose to move relative to the rest of their face; yet the two are undeniably connected. It is also frequently the case that cues are combined in complex ways. Although motion provides one cue to detachability, other cues such as coplanarity could be taken to imply that the surfaces are attached. The combination and weighting of conflicting cues is a complex topic, and it would be presumptuous to assume that any reduction in the strength of simultaneous contrast would be absolute.

The weakness of this particular experimental design is that it does not allow 'motion' conditions to be compared with 'no-motion' conditions. However, it is possible to compare different speeds of motion and I did investigate two speeds. Although this was primarily an attempt to discover the best frequency for performing the experiment, one might predict that faster motion would lead to a reduction in induction. However, my personal prediction is that it is unlikely that the slow-motion and fast-motion conditions would prove exactly equivalent, as the modulation frequency is different for both conditions. I have no basis to predict whether I would expect to see a strengthening or a reduction in simultaneous contrast, so even if the fast-motion condition causes a decrease in simultaneous contrast induction, it would be inadvisable to over-interpret this without further evidence.

Overview

Whilst in some respects this experiment was largely superseded by experiment 3.2 that allowed 'no-motion' and 'motion' conditions to be compared, I feel it still has some value because for most observers simultaneous contrast as measured by this method is extremely strong. Furthermore, this task was also subjectively easier and 'more natural' to carry out than experiment 3.2. Finally, it is primarily a 1st order motion stimulus rather than a 2nd order motion stimulus. For these reasons, I feel it is still worthwhile to include the results in this chapter.

Results

Experiment 3.1.1: LM-induction

In experiment 3.1.1a, we measure induction at two test speeds – slow (0.5 Hz) and fast (1 Hz). The LM contrast of the background is varied between +/-0.1 relative to the neutral point – in that portion immediately contiguous to the path of motion of the target.

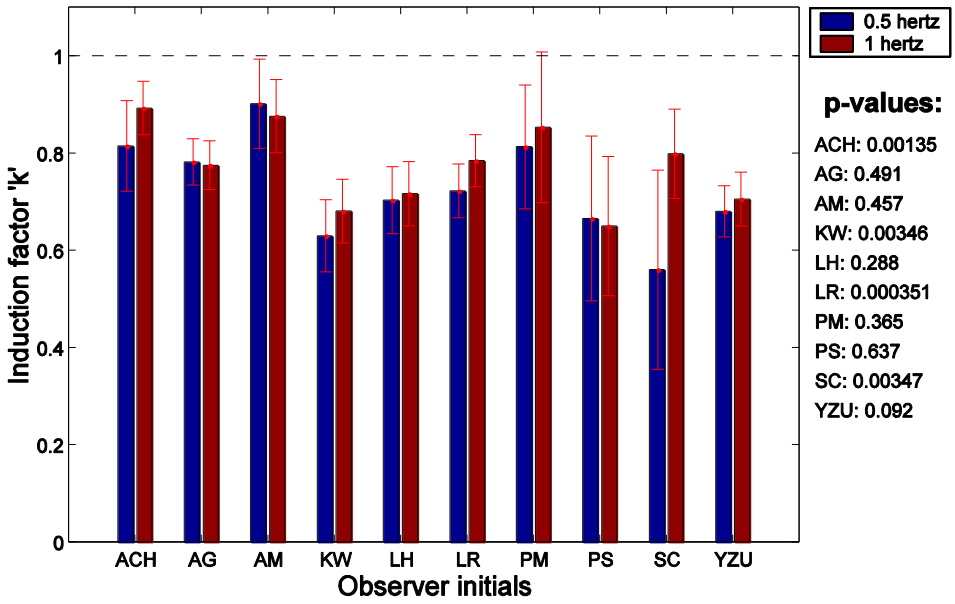


Figure 53: Bars show the strength of simultaneous induction for 10 observers, for the slow-motion / induction and the fast motion / induction conditions. The p-values reported were obtained from 1-way ANOVA applied to individual observers. For four observers, there was a significant difference ($p < 0.05$) between the 0.5 Hz (slow motion / induction) and 1Hz (fast motion / induction) conditions. In general, induction was stronger for faster motion.

Two-tailed p-values, determined for individual observers using ANOVA, are displayed to the right hand side of the graph. Error bars indicate 95% confidence intervals.

Figure 53 shows a graph of the results of this experiment. Table 16 shows the results of an ANOVA analysis performed on it, and table 17 shows descriptive statistics of the data, individualised for each observer.

The main result from experiment 3.1.1a is that induction is extremely strong for both the fast motion / induction and slow motion / induction conditions. When statistically significant differences were found between individual observer's settings for the slow motion and fast motion conditions, induction was always slightly stronger for the fast motion condition than for the slow motion.

The strength of induction elicited by these stimuli is higher than the value given by Zaidi et al (1991) for red-green induction of 40%-90% for two observers, and must compare favourably with the strongest values ever given for chromatic induction. Values for induction in experiment 2 were typically below 70%, so it seems clear that induction is not seriously reduced by the presence of simple motion cues. Indeed, the data suggest that induction may be stronger for the conditions with more pronounced relative motion, though these experiments cannot show whether this is due to the motion or to the different temporal frequency of the chromatic induction.

Source	Sum Sq.	d.f.	Singular?	Mean Sq.	F	Prob>F
Condition	0.186	1	0	0.186	16.539	0
Observer	3.212	9	0	0.357	31.685	0
Condition*Observer	0.387	9	0	0.043	3.815	0
Error	6.871	610	0	0.011		
Total	10.555	629	0			

Table 16: output from a 2-way unbalanced ANOVA of the raw data used to draw figure 57.

Table 16 shows the output of a 2-way ANOVA performed on the raw data generated in this experiment. Unsurprisingly, there are significant differences in the absolute levels of the settings made for each observer. However, although only a few observers demonstrated significantly different settings when compared on an individual basis, it appears that the consistent difference between the 0.5Hz and 1Hz conditions is statistically significant. Although a statistically significant interaction is found, it is much smaller in size than either of the main effects and can probably be ignored. Where a substantial effect is found (and the data do not appear obviously aberrant) the sign of the effect is always identical: induction is stronger for the faster condition.

Observer	*# trials*	*p value*	*Condition*	*Mean*	*STD*	*SEM*	*Kurtosis*	*Skew*	*'d'*
ACH	21	0.001	0.5 hertz	0.814	0.093	0.02	2.085	0.02	1.051151
			1 hertz	0.892	0.055	0.012	2.531	-0.384	
Trials per condition in each of 3 sessions: 1, 2, 18									
AG	55	0.491	0.5 hertz	0.782	0.048	0.006	2.504	-0.354	0.144038
			1 hertz	0.775	0.05	0.007	2.669	-0.666	
Trials per condition in each of 4 sessions: 3, 13, 20, 20									
AM	14	0.457	0.5 hertz	0.901	0.092	0.025	3.545	-0.78	0.309389
			1 hertz	0.875	0.075	0.019	1.865	0.021	
Trials per condition in 1 session: 15									
KW	32	0.003	0.5 hertz	0.63	0.075	0.013	1.662	0.128	0.723293
			1 hertz	0.68	0.065	0.012	1.978	-0.093	
Trials per condition in each of 4 sessions: 6, 20, 1, 6									
LH	58	0.288	0.5 hertz	0.703	0.069	0.009	2.402	0.314	0.198454
			1 hertz	0.716	0.066	0.009	2.338	-0.355	
Trials per condition in each of 3 sessions: 20, 20, 20									
LR	24	0	0.5 hertz	0.722	0.055	0.011	2.307	-0.444	1.139747
			1 hertz	0.784	0.054	0.011	1.901	-0.202	
Trials per condition in each of 5 sessions: 5, 5, 5, 5, 5									
PM	21	0.365	0.5 hertz	0.813	0.128	0.028	2.538	-0.769	0.283849
			1 hertz	0.853	0.155	0.034	2.966	-1.035	
Trials per condition in each of 4 sessions: 2, 10, 8, 2									

Observer	*# trials*	*p value*	*Condition*	*Mean*	*STD*	*SEM*	*Kurtosis*	*Skew*	*'d'*
PS	45	0.637	0.5 hertz	0.665	0.17	0.025	1.816	0.042	0.100201
			1 hertz	0.65	0.143	0.021	2.514	0.341	
Trials per condition in each of 3 sessions: 15, 15, 15									
SC	10	0.003	0.5 hertz	0.56	0.205	0.065	1.848	-0.568	1.609481
			1 hertz	0.798	0.092	0.029	1.853	-0.199	
Trials per condition in 1 session: 10									
YZU	22	0.092	0.5 hertz	0.68	0.052	0.011	2.975	-0.834	0.468959
			1 hertz	0.705	0.055	0.012	2.762	-0.262	
Trials per condition in each of 2 sessions: 2, 20									
Mean & Standard Deviations of Cohen's 'd'									
Mean: 0.602856			Standard deviation: 0.510942						
Overall homoscedasticity: 0.000 - the assumption of homoscedasticity was not met									

Table 17: descriptive statistics of the data from the moving-target condition; L-M axis chromatic induction

Experiment 3.1.2 - S-induction

Here, we repeat the previous experiment, using induction along the S-axis rather than the L-M axis. Stimulus luminances and spatial factors are as before, though the total S-cone contrast on the display ranges from +/-0.7 relative to the neutral 'whitepoint' defined earlier. Our predictions are identical to those in the previous experiment.

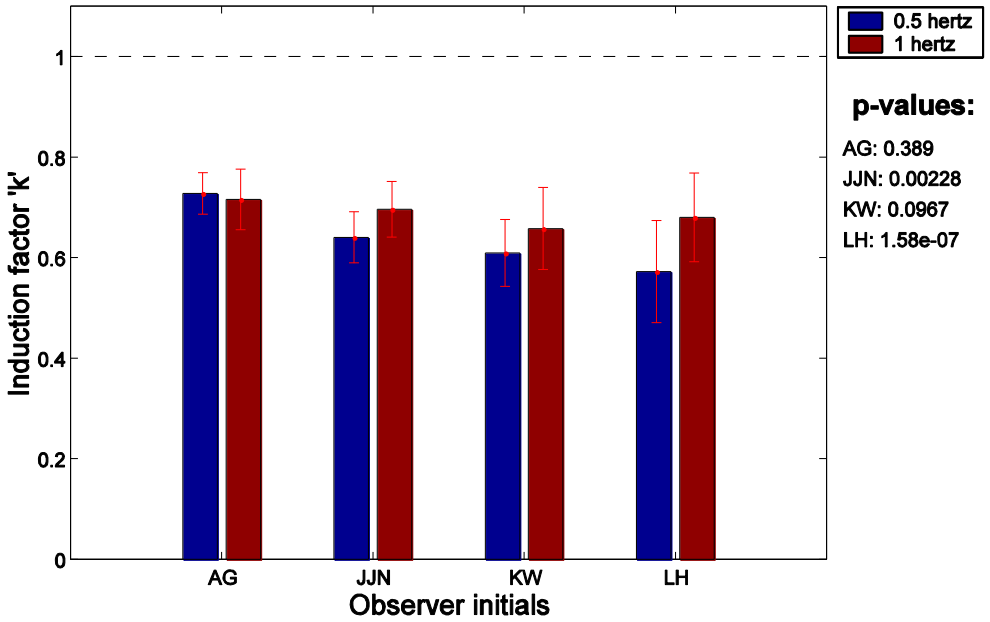


Figure 54: This graph shows the strength of induction along the S-axis for three observers. For two out of three observers, induction was significantly stronger for the fast-motion / induction condition than for the slow motion / induction condition ($p < 0.05$).

Two-tailed p-values, determined for individual observers using ANOVA, are displayed to the right hand side of the graph. Error bars indicate 95% confidence intervals.

Figure 54 shows our measurements for the strength of induction along the S-axis. The pattern of results is generally similar to that obtained for induction along the L-M axis, with most observers' data suggesting that induction is strongest for the fast-motion condition.

Table 18 and table 19 show the ANOVA table and descriptive statistics for this experiment. The pattern of results is very similar to those presented previously. Both the main effects are statistically significant, as is the interaction effect. Inter-observer differences account for the bulk of the variance, with differences between the slow and fast motion conditions, and the interaction term, accounting for much less of the variance.

Source	Sum Sq.	d.f.	Singular?	Mean Sq.	F	Prob>F
Condition	0.112	1	0	0.112	7.578	0.006
Observer	0.539	3	0	0.18	12.174	0
Condition*Observer	0.152	3	0	0.051	3.43	0.018
Error	3.454	234	0	0.015		
Total	4.353	241	0			

Table 18: output from a 2-way unbalanced ANOVA of the raw data used to draw figure 54

Observer	*# trials*	*p value*	*Condition*	*Mean*	*STD*	*SEM*	*Kurtosis*	*Skew*	*'d'*
AG	29	0.389	0.5 hertz	0.727	0.041	0.008	2.389	-0.577	0.231818
			1 hertz	0.716	0.06	0.011	2.643	0.133	
Trials per condition in each of 3 sessions: 10, 10, 10									
JJN									
	18	0.002	0.5 hertz	0.64	0.051	0.012	3.087	-0.764	1.048226
			1 hertz	0.696	0.055	0.013	3.045	0.13	
Trials per condition in each of 1 sessions: 20									
KW									
	14	0.097	0.5 hertz	0.609	0.067	0.018	4.239	0.696	0.654593
			1 hertz	0.658	0.081	0.022	2.09	-0.174	
Trials per condition in each of 1 sessions: 15									
LH									
	51	0	0.5 hertz	0.572	0.102	0.014	2.318	0.076	1.133721
			1 hertz	0.68	0.088	0.012	3.5	-0.202	
Trials per condition in each of 5 sessions: 10, 20, 15, 1, 10									
Overall homoscedasticity: 0.003 - the assumption of homoscedasticity was not met									
Mean & Standard Deviations of Cohen's 'd'									
Mean:	0.767090	Standard deviation: 0.413371							

Table 19: data used to draw figure 54

Experiment 3.1.3 - Brightness-induction

Here, we attempt to replicate our findings for chromatic induction, in the luminance axis. In the S and L-M axis experiments, luminance differences were used to ensure that the target did not disappear into the background around the neutral point, when both target and background had the same chromaticity. In this luminance version of the experiment, I used a chromatic contrast between the target and background to achieve the same effect. S-cone contrasts did not help localise the moving target, as isoluminant borders that are defined solely by S-cone contrasts are hard to detect. I therefore used an LM cone contrast of +0.1 (red) to localise the moving target against the background. In a control experiment, we measured the induction of a green target (LM-contrast -0.1) against the same, neutrally coloured background.

The background brightness is varied as much as possible: contrasts of +/-0.9 are used, relative to the reference colour. Stimulus sizes and mean luminances are as previously specified.

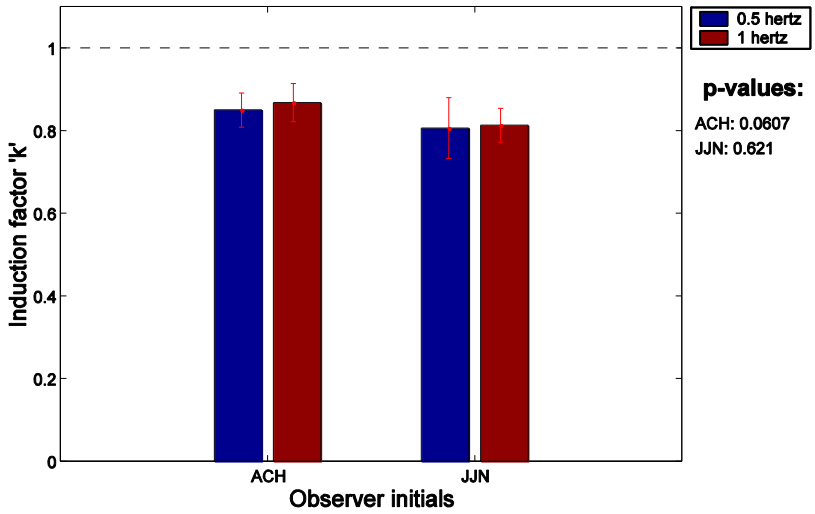
Prediction

Our prediction is that the pattern of data should be roughly similar to that for the chromatic axes.

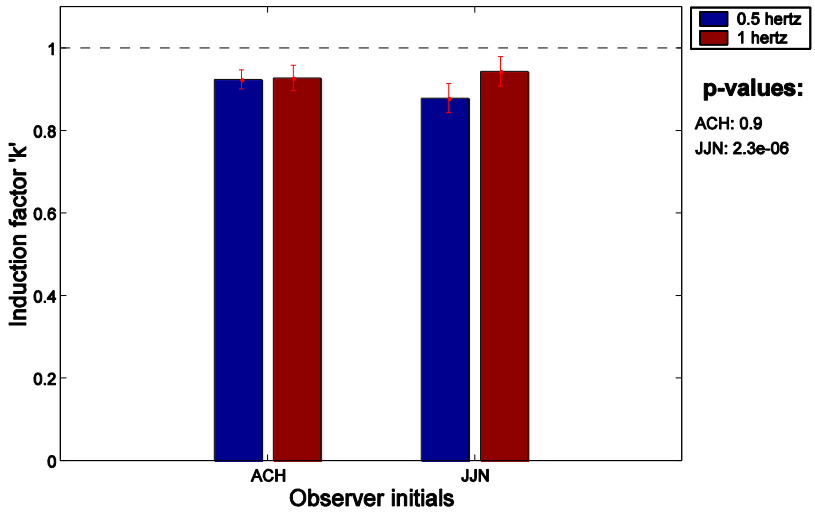
Results

Our results from this experiment are shown in figure 55a and figure 55b. Table 20 and table 21 show ANOVA tables for the experiments, and table 22 and table 23 show descriptive statistics of the raw data.

Induction remains strong, as for the chromatic axes. A statistically significant increase in the strength of induction is found for both conditions, though our interpretation of this should be tempered by the finding of a large interaction effect for the green target.



a) red



b) green

Figure 55: Induction along the luminance axis in a moving target square

a) for a red target square

b) for a green target square

Two-tailed p-values, determined for individual observers using ANOVA, are displayed to the right hand side of the graph. Error bars indicate 95% confidence intervals.

Note that induction is stronger for the condition with a green target square, than it is for the condition with a red target square. This pattern holds for all observers

Source	Sum Sq.	d.f.	Singular?	Mean Sq.	F	Prob>F
Condition	0.017	1	0	0.017	5.679	0.019
Observer	0.043	1	0	0.043	14.129	0
Condition*Observer	0	1	0	0	0.001	0.974
Error	0.383	126	0	0.003		
Total	0.447	129	0			

Table 20: results of ANOVA performed on the moving target experiment; luminance axis; red target The ANOVA was 2-way, unbalanced

Source	Sum Sq.	d.f.	Singular?	Mean Sq.	F	Prob>F
Condition	0.019	1	0	0.019	11.271	0.001
Observer	0.01	1	0	0.01	5.918	0.018
Condition*Observer	0.019	1	0	0.019	11.426	0.001
Error	0.111	66	0	0.002		
Total	0.165	69	0			

Table 21: results of ANOVA performed on the moving target experiment; luminance axis; green target The ANOVA was 2-way, unbalanced

Observer	# trials	p value	Condition	Mean	STD	SEM	Kurtosis	Skew	'd'
ACH	43	0.061	0.5 hertz	0.85	0.041	0.006	2.481	0.059	0.415586
			1 hertz	0.868	0.046	0.007	2.171	0.109	
Trials per condition in each of 1 sessions: 45									
JJN	18	0.621	0.5 hertz	0.806	0.073	0.016	2.201	-0.117	0.12262
			1 hertz	0.813	0.041	0.01	4.381	1.314	
Trials per condition in each of 1 sessions: 20									
Overall homoscedasticity: 0.314 - the assumption of homoscedasticity was met									
Mean & Standard Deviations of Cohen's 'd'									
Mean:	0.269103	Standard deviation:		0.207158					

Table 22: descriptive statistics for the moving target experiment; luminance axis; red target

Observer	# trials	p value	Condition	Mean	STD	SEM	Kurtosis	Skew	'd'
ACH	14	0.9	0.5 hertz	0.923	0.023	0.006	1.62	0.226	0.140261
			1 hertz	0.927	0.031	0.008	1.839	0.331	
Trials per condition in each of 1 sessions: 15									
JJN									
	19	0	0.5 hertz	0.878	0.035	0.008	1.998	-0.064	1.820396
			1 hertz	0.943	0.036	0.008	2.098	-0.192	
Trials per condition in each of 1 sessions: 20									
Overall homoscedasticity: 0.045 - the assumption of homoscedasticity was not met									
Mean & Standard Deviations of Cohen's 'd'									
Mean:	0.980328	Standard deviation:		1.188035					

Table 23: descriptive statistics for the moving target experiment; luminance axis; green target

Discussion

Taken together, the results from the moving-target experiments do not exclude an influence of motion cues on the strength of simultaneous contrast induction, but they are sufficient to show that motion segmentation does not strongly inhibit chromatic contrast induction. Indeed, induction measured here is considerably stronger than induction measured by other methods such as the 2AFC method employed by Thomas Wachtler et al. (2001), or by ourselves in previous experiments.

Performance on the basic task appears to be broadly similar in the LM, S and luminance axes. Contrary to predictions, increasing the speed of the motion does not appear to reduce the strength of the contrast induction effect. Whilst the results of individual observers are generally statistically insignificant, combined analysis with ANOVA confirms that the effect exists. However, it is still relatively small: in each case, between-observer variance is larger than between-condition variance.

Although the values of Cohen's 'd' are generally moderate to large, according to Cohen's standard definitions, the significance of any given value of 'd' does vary between fields of science. The figures reported here are far lower than some of the pseudo-'d' statistics calculated in the previous chapter. In view of this, and the fact that the between-observer variances are larger than the between-condition variances, it seems reasonable to conclude that effects described here are relatively insignificant in practical terms.

It would, of course, be possible to combine data from all three axes for simultaneous analysis. However, given that the design is already complicated by the unbalanced nature of the data, and in view of the fact that there are very different numbers of observers in each experiment, this seems inadvisable to me. We have already obtained significant findings for every effect for every condition, including interactions. Combining the analysis in this way would further test whether there was a difference between axes in the absolute extent of induction, and whether there was an interaction between the different axes. Although I would predict the answer to be 'yes' in both cases, neither of these hypotheses are currently of theoretical interest to me and are therefore best left unanalysed in order to reduce the experimentwise error rate.

The major disadvantage of this method is that we cannot compare motion conditions with no-motion conditions, because in the no-motion condition it is impossible to set a nulling match. Furthermore, the nulling frequency is confounded with the speed of the motion. These drawbacks mean that it is not possible to say whether motion affects the strength of simultaneous chromatic induction or not – my results from this experiment suggest that it may, but it is impossible to be sure because the speed of the motion is confounded with the frequency of the background change.

One natural control experiment that I would like to undertake in the future is to compare fast and slow rates of induction using a stationary target set against a uniform, stationary background whose entire colour was modulated. If induction is stronger for the fast-motion condition, this would suggest that the effect reported here is due to the frequency of induction rather than the increased speed of motion. However, ideally one would also seek to find a way of comparing motion and no-motion conditions directly. For this reason, I undertook the following set of experiments that examine induction against a moving background.

Experiment 3.2 - Are objects set against moving backgrounds subject to contrast effects?

Aims

This experiment builds on experiment 3.1, but deconfounds the relationship between the speed of motion and the frequency at which the colour of the inducing background is modulated. The stimuli used in this experiment allow comparisons between conditions when motion is either present or absent. The motion in these stimuli is also defined by texture and is therefore likely to favour second order motion-detection mechanisms over first-order motion mechanisms, thus expanding the scope of the experiment.

Methods

Stimulus parameters

Here, the target object remains stationary against a textured background that changes from red to green (or between two other specified endpoint colours) at a constant frequency. To avoid interactions, the background texture colours are chosen from a chromatic axis independent of the induced colour-shift. The textured surface in the background may be moved or left stationary, without interfering with other parameters of the experiment such as the frequency of the colour change of the inducing background.

As in experiment 3.1, the stimuli for this experiment filled the entire screen of the computer monitor, and subtended 30^o x 20^o of visual angle. For chromatic induction conditions, the background was equiluminant with a brightness of $30Cd/M^2$. To ensure the target never vanished into the background at those times when their mean chromaticities or brightnesses were identical, its colour was set 7% dimmer than that of the background (LM and S-axis chromatic induction conditions) or redder or greener than the background (brightness induction conditions).

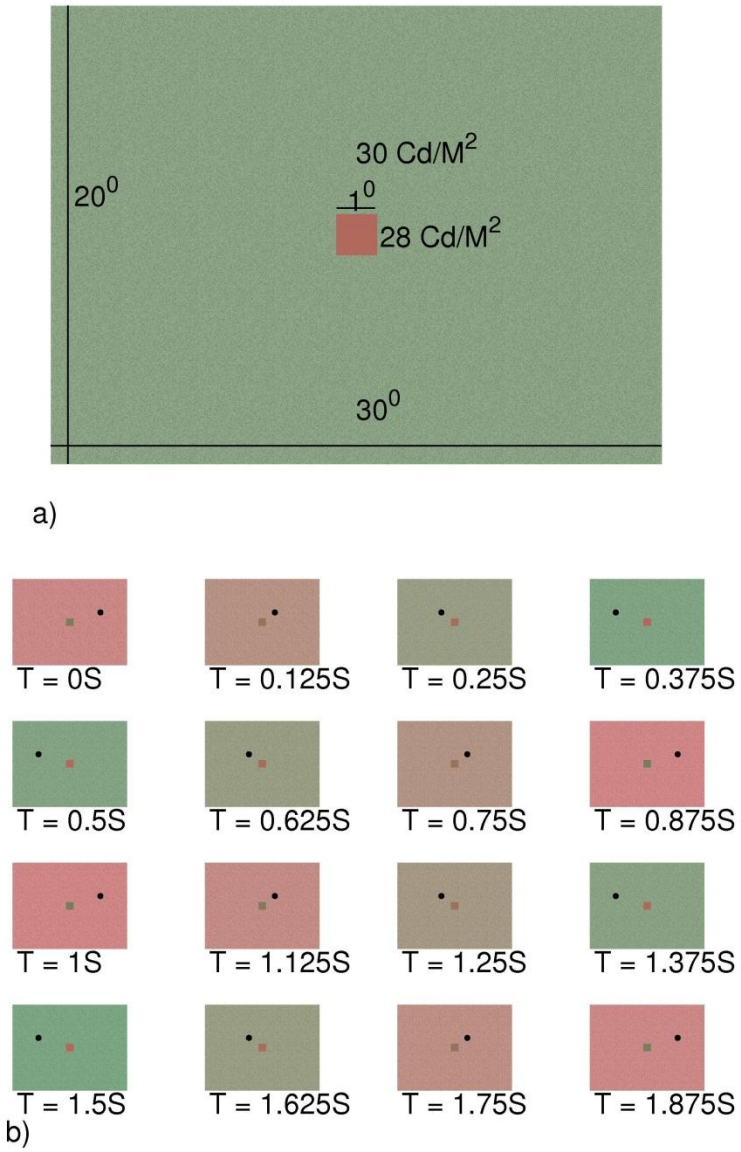


Figure 56: This figure illustrates (a) the basic stimulus parameters and (b) the change in colour and motion of the background.

a) The background subtends 30^o x 20^o of visual angle and has a brightness of $30Cd/M^2$. The brightness of the target square is set 7% dimmer to increase its contrast against the background. It subtends 1^o of visual angle.

b) Screenshots from the stimuli (1Hz condition) taken at constant intervals. In this experiment, the target square remains stationary in the center of the screen. The background is sinusoidally modulated along the L-M axis, and the textured background may move in phase with the colour change. To illustrate the background motion, a small black dot has been drawn at a fixed position on the background.

Experiment 2.2.3 showed that chromatic induction is tightly tuned to the cardinal opponent axes; therefore it is possible to manipulate the background (or target) along a cardinal axis without significantly disrupting induction along an independent axis. This finding is central to the technique used in this experiment, as some method of delineating the background position is necessary if its translation is to be perceived.

In the chromatic experiments (L-M and (L+M)-S induction) the background was covered with a random luminance texture with maximum contrasts of +/-0.25 relative to the reference luminance of 30Cd/M^2. In the luminance experiments, an L-M texture served to define the background position, with maximum cone-contrasts of +/-0.15 relative to the neutral chromaticity x=0.321, y=0.337.

The chromaticity of the background is modulated with a frequency of 1 Hz – causing induction equivalent to the 'fast motion' condition of the moving-target experiments. The chromatic axes and degree of modulation of the background are also identical to those previously employed.

The motion of the background takes the form of a sinusoidal oscillation along a horizontal plane, to minimise any effect similar to the waterfall illusion. Figure 56 illustrates this by defining a fixed point on the background, relative to the texture, showing its position during various parts of each cycle. However, inducing motion in this way means that 'Fast' and 'Slow' cannot be defined in terms of the velocity of the background displacement, as this is constantly changing. Instead, I define them in terms of the amplitude of the oscillation. If a point on the background moves with a period of 1 second and an amplitude of 8 degrees, it will clearly be moving faster than a point that moves with an amplitude of 1 degree, for all but the endpoints on its trajectory.

Stimulus verification

In a 136.19 second test run of the experimental program, comprising 11574 frames at a frame-rate of 84.99 Hz, only a single glitch was observed – and this was a single extra frame in duration.

Observer directions

Observers were not instructed to fixate on a single point on the screen; indeed this would have been difficult as no fixation point was provided. However, the 'natural' inclination was to fixate on the target square that remained in the center of the screen throughout the experiment.

Results

Experiment 3.2.1 – LM-induction

In this experiment, we modulate the background chromaticity along an equiluminant axis with L-M axis cone-contrasts of +/-0.1 relative to the reference background. We measure the strength of the resulting induction in the central target square at each of three conditions: no relative motion between the target and remote squares; motion with 1 degree of amplitude (slow), and motion with an amplitude of 8 degrees (fast).

Predictions

Here, we predict that induction should be weakened as the speed of background motion is increased.

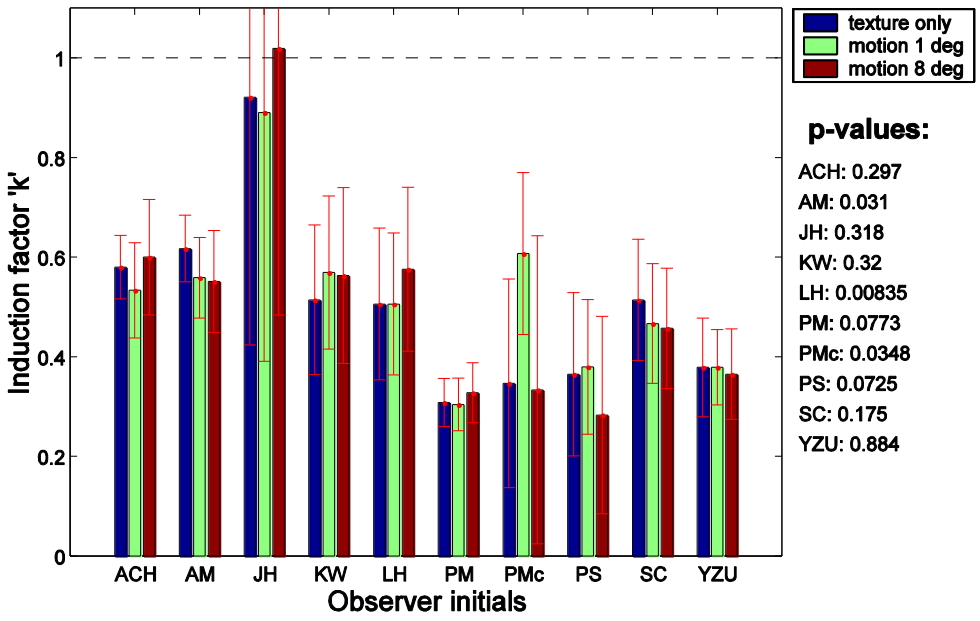


Figure 57: This graph shows the results for 10 observers, comparing the strength of induction for three conditions:

1) No-motion disparity between the target and the background.

2) Background motion amplitude of 1 degree of visual angle (target stationary)

3) Background motion amplitude of 8 degrees of visual angle (target stationary)

Two-tailed p-values, determined for individual observers using ANOVA, are displayed to the right hand side of the graph. Error bars indicate 95% confidence intervals.

For 8 out of 11 observers, there was no statistically significant difference between their settings for the three conditions. For the other three observers (ACH, LH and PMc) there does not appear to be any consistent pattern in the relationship between their settings for the three different conditions. Observer PMc found the task difficult to carry out. Observer JH's settings are somewhat higher than those of the other observers. A possible explanation for this is discussed in the text.

Figure 57 illustrates the results for the LM-axis version of this experiment. Only three observers showed a statistically significant difference between their results for the three conditions: AM, LH and PMc. Of these three observers, none showed a consistent reduction in the strength of simultaneous contrast for the motion-conditions relative to the no-motion conditions as expected. Arguably the data of observers SC is biased in this direction, but the trends were not statistically significant at the 5% level and must be viewed in the context of the results of the majority of observers that did not appear to show such a trend.

ACH showed no consistent difference between the motion and no-motion conditions; for observer LH induction appeared to become stronger as the motion-disparity between the background and the target increased. Observer JH obtained atypical results, mentioned later in the discussion, and observer PMc reported that it was very difficult to carry out the task and did not take part in further experiments.

Source	Sum Sq.	d.f.	Singular?	Mean Sq.	F	Prob>F
Condition	0.026	2	0	0.013	0.632	0.532
Observer	6.897	7	0	0.985	48.679	0
Condition*Observer	0.403	14	0	0.029	1.422	0.137
Error	14.511	717	0	0.02		
Total	21.876	740	0			

Table 24: results of a 2-way unbalanced ANOVA performed on the data for the 'moving texture' L-M axis induction condition

Table 24 shows the results of an ANOVA performed on the data, after atypical observers PMc and JH were excluded. As in the 'moving-center' experiments, there are clear inter-individual differences, however, inter-condition differences are not statistically significant, showing that either they are not present, or they are too small to be detected in this experiment. It is actually quite remarkable that no statistically significant differences are found between the three different conditions, in view of the large amount of data that was collected.

A point of interest is that the target patch can be adjusted either to appear as a neutrally coloured patch (the object of the experiment) or as a piece of neutral-density filter placed over the textured background. During training a few observers obtained values greater than '1' for the induction factor, indicating that to appear 'gray' the target square had to be either redder or greener than its immediate background. On questioning, one of these observers reported that they perceived the target square as being 'like a filter'. After being asked to view it as being 'like a piece of paper' he repeated the experiment, giving more reasonable results. I cannot contact observer JH to discuss his perception of the task, but his data are consistent with him carrying out the task in this way.

Observer	*# trials*	*p value*	*Condition*	*Mean*	*STD*	*SEM*	*Kurtosis*	*Skew*	*'d'*
ACH	9	0.297	texture only	0.58	0.064	0.02	2.622	0.159	
			motion 1 deg	0.533	0.096	0.032	2.071	-0.235	0.581903
			motion 8 deg	0.6	0.116	0.037	2.109	-0.451	0.223257
Trials per condition in each of 1 sessions: 10									
AM	19	0.031	texture only	0.617	0.067	0.015	1.848	-0.439	
			motion 1 deg	0.559	0.081	0.018	1.683	-0.467	0.795558
			motion 8 deg	0.551	0.102	0.023	2.538	0.61	0.78117
Trials per condition in each of 3 sessions: 7, 7, 7									
JH	73	0.318	texture only	0.92	0.497	0.057	2.882	-0.942	
			motion 1 deg	0.89	0.5	0.057	2.577	-0.873	0.060982
			motion 8 deg	1.019	0.535	0.063	3.653	-1.344	0.19068
Trials per condition in each of 5 sessions: 10, 20, 20, 20, 10									
KW	23	0.32	texture only	0.514	0.15	0.031	2.235	0.533	
			motion 1 deg	0.569	0.154	0.032	1.713	0.182	0.360541
			motion 8 deg	0.563	0.176	0.036	1.72	0.048	0.30013
Trials per condition in each of 5 sessions: 1, 5, 2, 5, 10									
LH	66	0.008	texture only	0.506	0.152	0.019	2.118	0.068	
			motion 1 deg	0.506	0.142	0.017	2.177	0.014	0.001012
			motion 8 deg	0.575	0.164	0.02	2.352	-0.43	0.44095
Trials per condition in each of 5 sessions: 10, 20, 20, 10, 10									

Observer	*# trials*	*p value*	*Condition*	*Mean*	*STD*	*SEM*	*Kurtosis*	*Skew*	*'d'*
PM	42	0.077	texture only	0.308	0.048	0.007	2.087	0.056	
			motion 1 deg	0.304	0.053	0.008	2.518	0.174	0.079425
			motion 8 deg	0.328	0.06	0.009	2.541	0.261	0.360862
Trials per condition in each of 3 sessions: 15, 15, 15									
PMc	9	0.035	texture only	0.347	0.209	0.07	1.781	-0.088	
			motion 1 deg	0.607	0.162	0.054	2.792	-0.99	1.401443
			motion 8 deg	0.334	0.309	0.103	2.349	-0.59	0.05052
Trials per condition in each of 1 sessions: 10									
PS	27	0.073	texture only	0.365	0.164	0.03	1.693	-0.136	
			motion 1 deg	0.38	0.135	0.026	2.268	0.568	0.100575
			motion 8 deg	0.283	0.198	0.037	2.219	-0.368	0.450523
Trials per condition in each of 3 sessions: 10, 10, 10									
SC	24	0.175	texture only	0.514	0.122	0.024	1.728	0.328	
			motion 1 deg	0.467	0.12	0.024	1.474	0.1	0.390118
			motion 8 deg	0.457	0.121	0.025	1.829	0.254	0.467954
Trials per condition in each of 3 sessions: 5, 15, 5									
YZU	21	0.884	texture only	0.378	0.099	0.022	2.741	-0.552	
			motion 1 deg	0.379	0.076	0.016	1.971	0.464	0.003084
			motion 8 deg	0.365	0.09	0.02	2.2	-0.026	0.141732
Trials per condition in each of 4 sessions: 2, 10, 8, 2									

Overall homoscedasticity: 0.000 - the assumption of homoscedasticity was not met

Mean & Standard Deviations of Cohen's 'd'

texture only:	**n/a**	
motion 1 deg:	**Mean: 0.377464**	**Standard dev.: 0.448**
motion 8 deg:	**Mean: 0.340778**	**Standard dev.:0.209**

Table 25: Data for the moving background experiment; L-M induction axis

Experiment 3.2.2 – S-induction

This experiment is an exact clone of the L-M moving-background experiment (3.2.1), except that the background is modulated along an S-cone stimulating axis. Our predictions are exactly as before: we expect relative motion to reduce the strength of the simultaneous contrast induction effect.

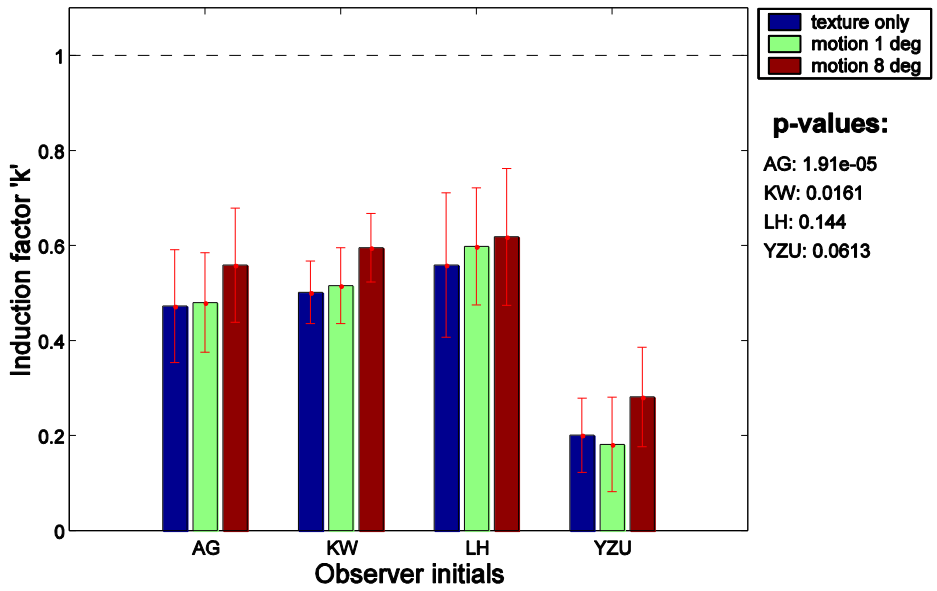


Figure 58: This graph shows the results for 4 observers, comparing the strength of induction for three conditions:

Two-tailed p-values, determined for individual observers using 2-way unbalanced ANOVA, are displayed to the right hand side of the graph. Error bars indicate 95% confidence intervals.

All four observers showed a trend towards increasing contrast induction with increasing motion-disparity between the moving background and fixed target square. For observers LH and YZU this trend was not statistically significant.

Figure 58 shows comparable results for induction along the S-axis. Table 26 shows an ANOVA table for the data and table 27 shows its corresponding descriptive statistics. For all observers, induction was stronger for the 8-degree motion amplitude condition (fast motion) but there was little difference between the no-motion and slow-motion conditions.

Observer YZU is notable by her much weaker general level of induction, and unsurprisingly between-observer differences are again highly statistically significant. However, despite the general trend towards higher levels of induction for higher speeds of relative motion, the main effect between conditions does not appear to be significant, even though it is for observers KW and AG when analysed individually. This could be related to the strong interaction effect, the low power of type-III sum of squares ANOVA or simply the pattern of results that makes it difficult for ANOVA to detect a difference.

Source	Sum Sq.	d.f.	Singular?	Mean Sq.	F	Prob>F
Condition	0.193	2	0	0.097	2.807	0.062
Observer	2.808	3	0	0.936	27.198	0
Condition*Observer	0.551	6	0	0.092	2.668	0.015
Error	14.35	417	0	0.034		
Total	17.996	428	0			

Table 26: 2-way unbalanced ANOVA for the data shown in figure 58

Observer	*# trials*	*p value*	*Condition*	*Mean*	*STD*	*SEM*	*Kurtosis*	*Skew*	*'d'*
AG	61	0	texture only	0.472	0.119	0.015	2.732	-0.366	
			motion 1 deg	0.48	0.105	0.013	1.903	0.158	0.067678
			motion 8 deg	0.558	0.12	0.015	2.343	-0.181	0.723263
Trials per condition in each of 5 sessions: 10, 15, 15, 15, 10									
KW	10	0.016	texture only	0.501	0.066	0.021	2.359	0.581	
			motion 1 deg	0.515	0.08	0.025	1.569	0.122	0.195023
			motion 8 deg	0.595	0.072	0.023	2.5	-0.811	1.359273
Trials per condition in each of 1 sessions: 10									
LH	53	0.144	texture only	0.559	0.152	0.02	1.911	-0.078	
			motion 1 deg	0.598	0.123	0.016	1.935	-0.152	0.284733
			motion 8 deg	0.618	0.144	0.02	3.551	-0.439	0.399637
Trials per condition in each of 4 sessions: 15, 14, 14, 15									
YZU	10	0.061	texture only	0.201	0.078	0.025	2.173	-0.418	
			motion 1 deg	0.181	0.099	0.031	2.678	-0.56	0.216486
			motion 8 deg	0.281	0.105	0.033	2.449	-0.084	0.878307
Trials per condition in each of 1 sessions: 10									

Overall homoscedasticity:	**0.001 - the assumption of homoscedasticity was not met**			
Mean & Standard Deviations of Cohen's 'd'				
texture only:	**n/a**			
			Standard	
motion 1 deg:	**Mean:**	**0.19098**	**deviation:**	**0.090664**
			Standard	
motion 8 deg:	**Mean:**	**0.84012**	**deviation:**	**0.399441**

Table 27: descriptive statistics for the S-axis induction condition; moving background texture

Experiment 3.2.3 – Luminance-induction

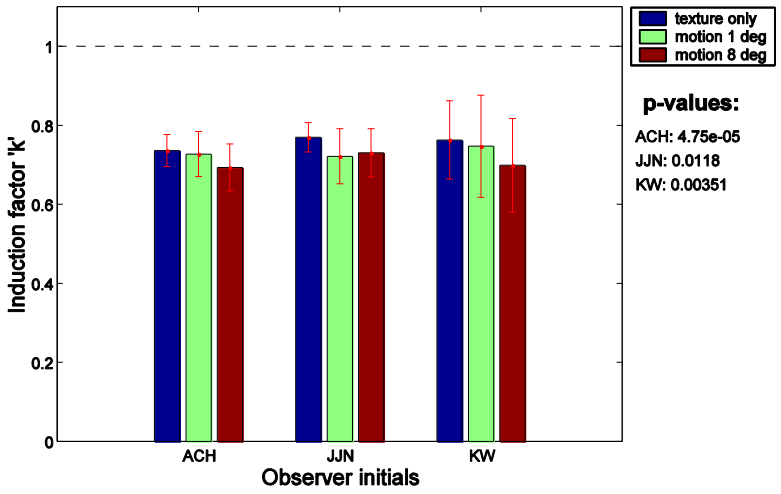
Figure 59a and b show values for luminance induction against stationary or moving backgrounds. Table 28 and table 29 show ANOVA tables for the experiment, and table 30 and table 31 show the corresponding descriptive statistics. Again, the absolute values obtained for induction were quite high for all conditions. Although table 28 shows that there is a significant difference between conditions for a red target, in absolute terms this is not found in table 29 (green target), and in absolute terms the effect is still very small. In contrast to findings for the chromatic axes, they differences here *were* in the direction of decreasing contrast induction with increasing relative motion, which is consistent with my predictions. However, the smallness and inconsistency of the effect argues against it being meaningful.

Source	Sum Sq.	d.f.	Singular?	Mean Sq.	F	Prob>F
Condition	0.127	2	0	0.063	6.284	0.002
Observer	0.022	2	0	0.011	1.069	0.344
Condition*Observer	0.025	4	0	0.006	0.63	0.641
Error	3.685	366	0	0.01		
Total	3.901	374	0			

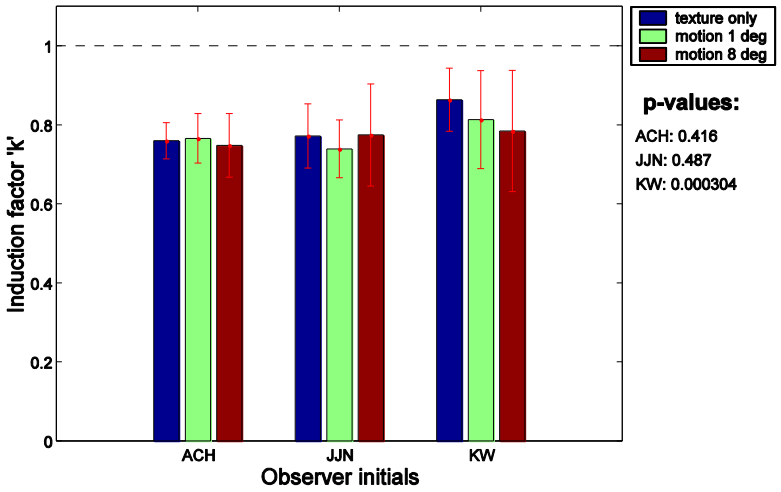
Table 28: 2-way unbalanced ANOVA table for the moving background experiment; luminance condition; red target

Source	Sum Sq.	d.f.	Singular?	Mean Sq.	F	Prob>F
Condition	0.051	2	0	0.026	2.091	0.125
Observer	0.213	2	0	0.106	8.728	0
Condition*Observer	0.094	4	0	0.024	1.93	0.105
Error	3.807	312	0	0.012		
Total	4.193	320	0			

Table 29: 2-way unbalanced ANOVA table for the moving background experiment; luminance condition; green target



a)



b)

Figure 59a and b: Luminance induction in red and green target squares

a) Red target square (contrast 0.1 relative to neutral)

For all observers, induction is weaker in the 8-degree motion condition than in the no-motion condition, though this difference is small in real terms.

b) Green target square (contrast -0.1 relative to neutral)

Only observer KW shows a statistically significant difference between the no-motion condition (texture-only) and the moving background conditions. The pattern of his data in figures a) and b) is quite similar.

For all observers, induction is slightly stronger for the condition with a green target square, than it is for the condition with a red target-square.

Two-tailed p-values, determined for individual observers using ANOVA, are displayed to the right hand side of the graph. Error bars indicate 95% confidence intervals.

Observer	*# trials*	*p value*	*Condition*	*Mean*	*STD*	*SEM*	*Kurtosis*	*Skew*	*'d'*
ACH	58	0	texture only	0.736	0.04	0.005	2.018	-0.04	
			motion 1 deg	0.727	0.057	0.007	2.405	0.048	0.174664
			motion 8 deg	0.693	0.06	0.008	1.89	-0.126	0.854053
Trials per condition in each of 3 sessions: 15, 15, 15									
JJN	19	0.012	texture only	0.769	0.037	0.009	3.735	0.886	
			motion 1 deg	0.721	0.07	0.016	1.919	0.197	0.89954
			motion 8 deg	0.73	0.061	0.014	2.831	-0.105	0.798769
Trials per condition in each of 1 sessions: 20									
KW	74	0.004	texture only	0.763	0.099	0.011	2.019	0.113	
			motion 1 deg	0.747	0.129	0.015	2.032	-0.169	0.1424
			motion 8 deg	0.699	0.119	0.014	2.388	-0.129	0.591723
Trials per condition in each of 3 sessions: 20, 20, 20									

Overall homoscedasticity: 0.000 - the assumption of homoscedasticity was not met

Mean & Standard Deviations of Cohen's 'd'

texture only: n/a

motion 1 deg: Mean: 0.405535 Standard deviation: 0.428125

motion 8 deg: Mean: 0.748182 Standard deviation: 0.138288

Table 30: descriptive statistics of the data for the moving-background, luminance induction experiment (red target square) – as used to draw graph 11a.

Observer	*# trials*	*p value*	*Condition*	*Mean*	*STD*	*SEM*	*Kurtosis*	*Skew*	*'d'*
ACH	56	0.416	texture only	0.759	0.046	0.006	2.836	0.513	
			motion 1 deg	0.765	0.063	0.008	2.374	0.148	0.111073
			motion 8 deg	0.748	0.081	0.01	1.876	0.197	0.181857
Trials per condition in each of 4 sessions: 8, 7, 8, 15									
JJN	19	0.487	texture only	0.772	0.081	0.018	2.019	0.395	
			motion 1 deg	0.739	0.073	0.017	1.731	-0.227	0.425973
			motion 8 deg	0.774	0.129	0.029	2.594	-0.332	0.022892
Trials per condition in each of 1 sessions: 20									
KW	78	0	texture only	0.864	0.08	0.009	1.863	0.056	
			motion 1 deg	0.813	0.124	0.014	2.564	-0.206	0.495943
			motion 8 deg	0.784	0.153	0.017	2.261	0.072	0.682835
Trials per condition in each of 4 sessions: 20, 20, 20, 1									

Overall homoscedasticity: 0.000 - the assumption of homoscedasticity was not met

Mean & Standard Deviations of Cohen's 'd'

texture only:	n/a	
motion 1 deg:	Mean: 0.344330	Standard deviation: 0.205013
motion 8 deg:	Mean: 0.295861	Standard deviation: 0.344425

Table 31: descriptive statistics of of the data for the moving-background, luminance induction experiment (green target square) – as used to draw graph 11b.

Discussion

The results of this set of experiments with a moving background appear to be somewhat contradictory. We obtain many individual results that are strongly statistically significant, but overall they do not seem to amount to a clear pattern. Our findings, summarised, are as follows:

For the LM-axis, three of ten observers show statistically significant results, but in each case the pattern of results is different, and overall there is no effect.

For the S-axis, two of four observers show similar and statistically significant results, but overall there is no statistically significant effect.

For the luminance axis, there is a statistically significant effect in the direction predicted (motion reduces contrast induction), but this is true only for the red-target condition.

As in the moving-target experiment, our results generally fail to confirm our prediction that induction should be strongly reduced. The between-condition variances generally appeared to be even smaller than in the previous experiment, and were rarely found to be statistically significant. This probably reflects the fact that the induction frequency no longer changed between conditions, and weakly supports the notion that the effect of motion/induction speed in experiment 3.1 was probably primarily due to the frequency of induction rather than the speed of motion.

Given that the stimuli in this experiment will have stimulated the second-order motion mechanisms, and the stimuli in experiment 3.1 favoured first-order motion mechanisms, when taken together our results suggest that neither form of motion is able to block simultaneous contrast induction. Given that almost all moving objects in the natural world will activate either first or second order motion mechanisms, it seems unlikely that differential motion plays a strong part in the modulation of simultaneous chromatic contrast induction.

On the other hand, we have obtained several statistically significant results that, although small, do in some cases appear too ‘statistically significant’ to be ascribed to type-I error. One possibility is that motion genuinely increases S-cone induction but reduces luminance induction, though I do not immediately see why this should be. A further possibility is that motion acts indirectly on induction, perhaps by stimulating eye movements. More assiduous observers may be less susceptible to these effects. It’s also important to remember that observers are not blinded to the experiment, so small subconscious biases could also account for these results.

Experiment 3.3 – Can differences in perceived depth disrupt simultaneous contrast?

Aims

In this experiment, we cause the target object to appear in the same plane as the inducing background, in front of it, or behind it. We compare the strength of simultaneous contrast induction in each case.

Predictions

In this experiment, we would predict any depth differences between the target and its background to reduce the strength of simultaneous contrast induction; that is, induction should be strongest for the zero-disparity condition.

Methods

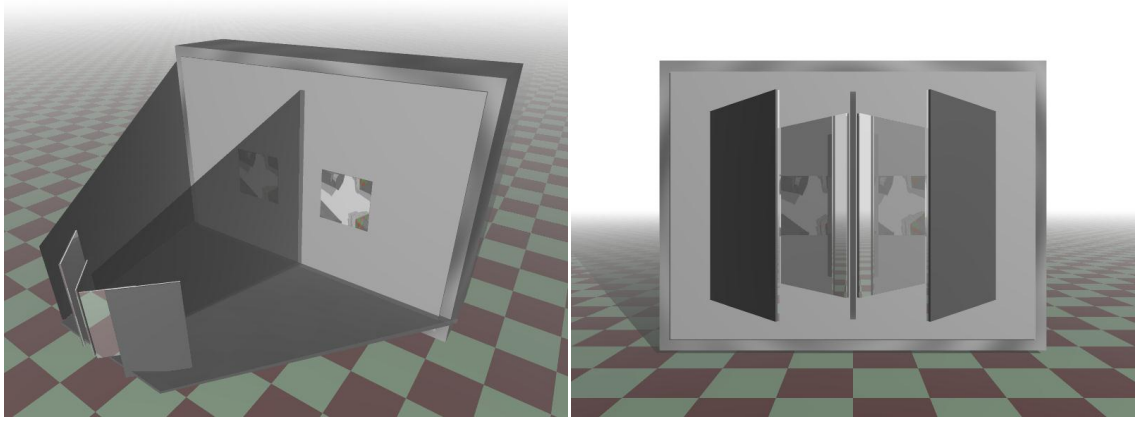
Methods of eliciting differences in perceived depth

A number of cues can provide information about the absolute distance of an object from an observer, or of relative depth differences between two or more different objects. These cues include convergence and binocular disparity, which can provide depth information about objects up to around 6 metres distant from the observer, as well as other cues such as motion-parallax and size that may be useful over larger distances. Convergence and disparity cues to depth are obtained by comparing the areas in the different images formed in each eye for absolute or relative offsets. Size cues depend on prior knowledge – for example most adults are between 1.5 and 2 metres tall, so we know that a person who appears tiny is likely to be distant rather than short in stature. Parallax cues depend on motion – when looking out of a train window, trees close to the track appear to pass quickly but mountains distant from the observer appear stationary over short periods of time. To provide parallax cues, a cat may move its head up and down as an aid to judging distance, prior to taking a risky leap.

Unfortunately it is difficult to recreate most of these cues effectively or completely realistically on a computer monitor. For example, to recreate the cue of motion parallax an experimenter would have to monitor the position of the subject's head very accurately and modify the display in response to any movement. Likewise we could attempt to simulate blurring in surfaces outside of the horopter, but to do this we would need to monitor the subject's 3d point of fixation using a binocular eye-tracker, and measure his or her pupil diameter in order to estimate the correct depth-of-field.

A further caveat is that not all people have good vision in both eyes, and even those that do may not be sensitive to binocular depth cues. In the region of 10% of Americans tested have a deficit of their stereoscopic vision, and this is particularly true for those who suffered from amblyopia or strabismus during childhood. As with the anomalies of colour-vision, affected individuals are often not aware of their condition as they can lean on other cues such as motion parallax or shadowing that can compensate most of the time.

Wheatstone stereoscope



a) b)

Figure 60: Cutaway diagrams of a Wheatstone stereoscope

The Wheatstone stereoscope is composed of four mirrors, and baffles arranged so that the right and left eyes view two adjacent pictures, respectively.

a) An oblique view of the four mirrors used to split the image, and of the baffles used to prevent the right and left eyes from seeing the image intended for the other eye. Two similar images are presented on each side of the screen and in practice; observers must adjust the angle of the mirrors until both images may be fused comfortably.

b) A frontal view of the apparatus, showing how the two images on each half of the monitor are reflected through the inner and outer sets of viewing mirrors. At this viewing distance, it is possible to see both images simultaneously. At closer viewing distances, each eye can only see the image that is intended for it.

Wheatstone apparatus

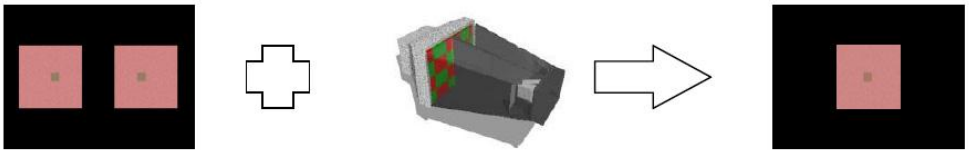


Figure 61: using the computer and Wheatstone stereoscope to display dichoptic images

The left and right eyes each see separate halves of the computer screen – the partition in the Wheatstone viewer prevents each eye from being able to see the stimulus intended for the other eye.

The two views are combined in the brain to form a single stable percept of a small target square set against a larger textured background. In the example shown above, the small square would appear to be floating in front of the large square.

Stimuli

Due to the use of the Wheatstone apparatus, a smaller stimulus area was available for these experiments than for experiments 3.1 and 3.2, and each background square in this experiment subtended only 6^o by 6^o. However, the brightness, colours and textures used for the background were the same as for experiment 3.2, and the target dimensions (1^o square) were also identical.

The Wheatstone apparatus has to be adjusted individually for each observer, to account for their differing inter-ocular distances, so prior to starting each experiment the observers adjusted the mirror settings until they could fuse both background squares comfortably. Although each observer viewed the stimuli from a fixed distance of 67 cm, the convergence distance may have differed slightly for each observer, or even between experimental sessions for the same observer. However, the purpose of the experiment is to examine whether displacements in depth of the target square relative to the background modified the strength of simultaneous contrast. Fixed displacements of $+0.5^{o}$ (near condition), -0.5^{o} (far condition) and 0^{o} (same plane) relative to the background were used to cause the target square to appear nearer to the observer than the background, behind the background, or in the same depth plane as the background. These three disparities were interleaved within each session of observing, but the target square remained exactly the same physical size in each condition. This meant that the target square's perceived size may not have been constant throughout each experimental session as when binocular disparity cues indicated that it was closer, it appeared physically smaller than when disparity cues indicated that it was further away. However, the size of the target and background squares on the retina remained constant throughout the experiment. The exact displacements along the Z-axis are dependent on the perceived distance of the background plane from the observer, which may have varied with the convergence angle. It is therefore not possible to provide equivalent measurements in other units such as centimetres of depth-difference between the plane of the surround and the uniform targets.

These disparities were chosen by a process of trial-and-error, and this was the maximum disparity that was easy and comfortable to fuse for most observers. In each trial, the depth-cues to the position of the target relative to the textured background remained constant, another ploy to encourage stable fusion (in pilot experiments I found that it was hard to track a target that moved in depth over the course of the experiment).

Exclusion criteria

Observers who found it difficult to fuse the stimuli or who were unable to perceive depth when using the stereoscope, were not asked to carry out this experiment. A control experiment was run where prospective observers were asked to state whether the target square was closer than, further than or in the same plane as the background. If they were not able to accomplish this task consistently, they were excluded from the experiment.

Results

This experiment was performed only along the LM axis, for three observers.

Figure 62 shows that induction is equally strong for all conditions for most observers ($p > 0.05$). Although it is possible that observers may show small differences between the various conditions, there does not seem to be any fixed pattern, and induction appears to remain strong throughout the course of the experiment. We therefore find no evidence that binocular disparity can disrupt simultaneous contrast.

L. Maloney questioned (personal communication) whether in experiment 3.2 (moving texture) some observers might perceive the target as being underneath a translucent background. If so, I might effectively be investigating how we perceive the colours of objects viewed through coloured filters, rather than the influence of simultaneous contrast over perceived colour appearance. But the results from this related experiment seem to counter this possibility – at least for a majority of observers.

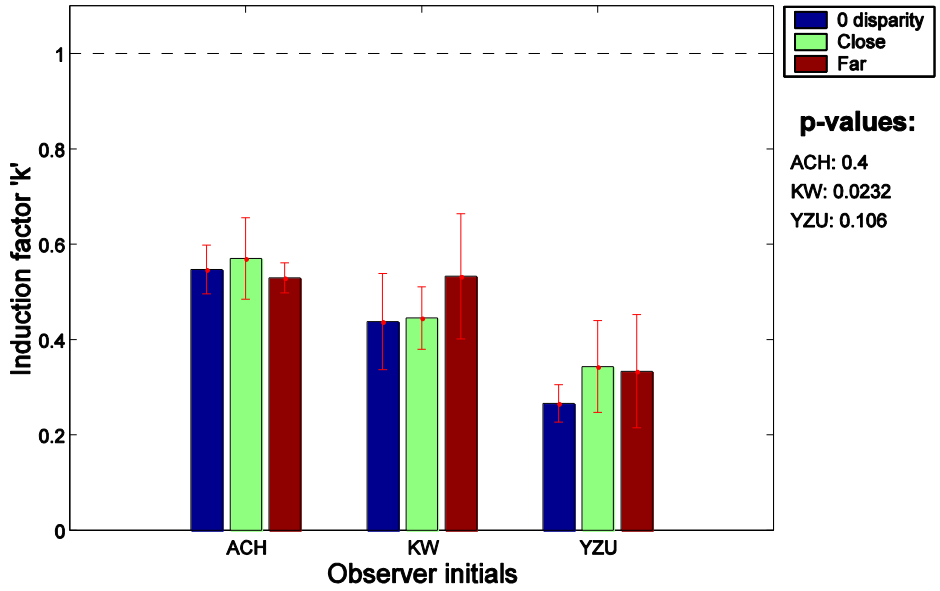


Figure 62: Induction strength for three observers for three conditions:

1) The target square and background set against the same plane

2) The target square placed closer to the observer than the inducing background

3) The target square viewed as if through a window in the inducing background.

Two-tailed p-values, determined for individual observers using ANOVA, are displayed to the right hand side of the graph. Error bars indicate 95% confidence intervals.

None of the observers show a statistically significant difference between any of the conditions, and there is no consistent pattern between the three observers.

In one condition (far condition) the target appears to be further away than the inducing background, whereas in the no-disparity experiment the target appears to be coplanar with the inducing background, and in the third condition, the target square is perceived as being closer than the inducing background. An observer should be more likely to perceive the target square as being viewed through a filter if it is set behind or in the same plane as the inducing background, than if binocular disparity suggests that it is closer than the inducing background. However, for two out of three observers, there is no significant difference between the settings made when the target square is perceived as being near or far from the inducing background. For observer KW, there was a small difference that approached statistical significance, but this result that affected the distant setting only and there was no difference between the no-disparity and the near conditions.

Source	Sum Sq.	d.f.	Singular?	Mean Sq.	F	Prob>F
Condition	0.114	2	0	0.057	2.349	0.1
Observer	1.167	3	0	0.389	16.025	0
Condition*Observer	0.405	6	0	0.067	2.779	0.014
Error	3.131	129	0	0.024		
Total	4.82	140	0			

Table 32: results of a 2-way unbalanced ANOVA performed on data from the disparity experiment

Table 32 shows the ANOVA output analysing the data from the disparity experiment. There was a statistically significant interaction effect which was larger than any putative effect of condition. However, inter-individual differences in the strength of induction were apparent, as ever.

Observer	*# trials*	*p value*	*Condition*	*Mean*	*STD*	*SEM*	*Kurtosis*	*Skew*	*'d'*
			0						
ACH	9	0.4	disparity	0.547	0.051	0.016	1.98	0.405	
			Close	0.57	0.085	0.027	2.397	0.024	0.337854
			Far	0.529	0.031	0.01	1.687	-0.147	0.42502
Trials per condition in each of 1 sessions: 10									
			0						
KW	19	0.023	disparity	0.437	0.101	0.023	7.946	-1.848	
			Close	0.445	0.066	0.015	1.702	-0.25	0.093018
			Far	0.532	0.131	0.029	4.627	1.368	0.816786
Trials per condition in each of 3 sessions: 1, 10, 10									
			0						
YZU	11	0.106	disparity	0.266	0.039	0.012	2.086	0.793	
			Close	0.343	0.096	0.028	2.143	0.304	1.140421
			Far	0.333	0.119	0.036	2.096	-0.123	0.85703
Trials per condition in each of 3 sessions: 2, 8, 2									
Overall homoscedasticity: 0.279 - the assumption of homoscedasticity was met									
Mean & Standard Deviations of Cohen's 'd'									
0 disparity: n/a									
Close: Mean: 0.523764 Standard deviation: 0.547891									
Far: Mean: 0.699612 Standard deviation: 0.238653									

Table 33: descriptive statistics for the binocular disparity experiment

Discussion and conclusions

We obtained broadly similar results from all three experiments (moving target, moving background, binocular disparity). In all cases, induction seemed to be quite strong along all the cardinal axes, but data from several observers suggests that both motion and disparity may have some small but real influence on the strength of simultaneous induction. This may be either an enhancement or a suppressive effect depending on the observer and the condition.

Overall, these results should be taken as strongly disconfirming my working hypothesis that motion and depth contrasts should strongly reduce the strength of simultaneous contrast. But several of them did reach statistical significance. Is it valid to simply dismiss these statistically significant findings, or can they be interpreted in the 'mode of exploration' – i.e. can we use them as a basis for a phenomenological description of the influence of motion on contrast induction? For example, if we had found that motion *increased* induction in all conditions, and this result was statistically unambiguous and of clear practical significance, it would be proper to try and think of possible explanations for the finding despite the fact that it was not predicted.

The main factor that leads me to believe that the differences that I have found are probably not functionally significant, is their inconsistency. Although it may not be surprising to find that there are different patterns of results between axes (even if the nature of these patterns was not predicted), any finding of qualitative differences between observers is less easy to explain. I find it hard to reconcile this observation with an important role of simultaneous contrast in influencing colour perception: though of course it is possible that we each have an independent 'algorithm' for colour constancy, this would be unparsimonious.

Furthermore, the influence of motion or depth cues on intra-observer differences for each cardinal axis is generally quite small. Clearly the importance of large effects in determining real-world colour-perception is likely to be greater than that of diminutive effects, but the threshold at which an effect ceases to be 'functionally significant' must be at least partly arbitrary. My personal opinion is that effects of the magnitude observed here are unlikely to be important outside of the laboratory.

Could the possible between-observer or between-axis differences suggested by experiment 3.2 have a theoretical significance, even if they are not functionally important? This is certainly possible, but I do not currently have any framework in which to interpret them. Furthermore, when confronted with unanticipated or seemingly contradictory results, the correct course of action is replication rather than over-interpretation.

Chapter 7 - Conclusions and Potential Future Areas of Research

In the introduction, I made a number of predictions regarding the conditions under which simultaneous contrast should act, if it is to improve colour-constancy. I expected that contrast would be weakened when segmentation cues indicated that the target object and surround were separate from each other. It is of course possible that our premises were invalid – there is no data currently available that I am aware of that would allow us to test them. However, I hope that as hyperspectral imaging becomes more ubiquitous and we learn more about the statistics of natural images, they may be put to the test.

I investigated motion, depth and texture cues, and most of our findings contradicted our predictions, with the exception of some of the experiments investigating textured surfaces. Neither motion cues nor depth cues seemed to appreciably influence simultaneous contrast, however texture either in the inducing background or in the target object, appeared to reduce the strength of simultaneous contrast induction very significantly. Even here, the experimental results are not as I predicted; the degree of asymmetry between the effects of texture in the surround and texture in the target square suggest a low-level effect based on the properties of the texture, rather than on any figure-ground segmentation.

At present none of my individual results show whether the reduction in simultaneous contrast induction was due to segmentation or other properties of the stimuli, however my results did suggest that induction involves early mechanisms in visual processing and provide strong evidence against my prediction that all segmentation cues should influence simultaneous contrast induction. The fact that induction seems to involve an early mechanism in visual processing does not preclude the possibility that mid-level functions, such as segmentation, could be involved via feedback pathways. However to me it seems surprising that one cue might influence simultaneous contrast so strongly, yet other cues have little or no effect – if the cues are processed interdependently as I assumed that segmentation cues would probably be. My view is that both texture-processing (at least as pertinent to my stimuli) and simultaneous contrast induction are likely to take place very early in the visual system, and probably work independently of higher level representations of the visual world.

Future research questions

Distinguishing between different 'simultaneous contrast' mechanisms

My data clearly show that for simple stimulus configurations, stereoscopic depth and motion segmentation have very little effect on the simultaneous contrast effect. Yet other experimenters have created stimuli where depth is an extremely powerful determinant of surface brightness perception – and as discussed in my introduction, all of these have at some time or another been classed under the single banner of 'simultaneous contrast'.

There is general agreement, at least amongst physiologists, that colour perception is determined by a number of mechanisms sited at different levels of the visual system, beginning with the three cone-types in the retina. In my view it is important to distinguish between different effects in the study of the visual system, not only in terms of physiology, but also in terms of computational processes and psychophysical and perceptual effects. For example, it is likely that 'simultaneous contrast' as elicited by my stimuli, is processed in an utterly different part of the brain to contrast as elicited by Alan Gilchrist's demonstrations of illumination frameworks (Gilchrist 1977). I have shown that my stimuli are processed at a basic level of the visual system, but the perceived lightness of Alan Gilchrist's must be influenced by more complex processes such as depth perception.

My personal goal in studying the visual system is to relate function to physiology; to this aim it is clearly desirable to differentiate between effects that appear superficially similar but probably involve different physiological mechanisms. Whilst there are alternative and equally valid approaches to studying the visual system, in my view most would benefit from a clear framework of how different visual effects and illusions relate to one another. For example, the stimuli used to elicit the White illusion bear a passing resemblance to some of my stimuli. Could their effects be in any way related? Annette Werner found that we adapt more quickly to articulated backgrounds than to uniform backgrounds (Werner, Sharpe et al. 2000), which some consider at odds with my result that induction is stronger for uniform backgrounds than for articulated ones. Are our results truly in conflict or is it inappropriate to compare them because we are studying different effects, as I believe?

A taxonomy of visual illusions is an ambitious and probably unattainable goal, and I suspect that it may prove difficult or impossible to 'isolate' and distinguish between mechanisms that putatively influence our perceptions of more complex stimuli. Indeed, in my view it is highly unlikely that our perception of complex stimuli will ever be explained by single, isolated mechanisms. Bloj et al. have shown that the perceived geometric arrangement of two surfaces can influence the colours we perceive the surfaces to be, as with some arrangements we correct for inter-reflections and for other arrangements we know that such inter-reflections cannot occur (Bloj, Kersten et al. 1999). The parsimonious assumption is that such a complex task that draws on our perceptions of geometry, depth and colour, is carried out by the visual system working as a cohesive whole rather than by a single highly advanced mechanism working in isolation.

However, I am optimistic that we may be able to differentiate between 'chromatic adaptation' and 'chromatic induction' in the future. In my own thinking, I view 'induction' as an effect that probably occurs all the time, but which starts to act within a fraction of a second. 'Adaptation' to a background colour can occur fairly quickly and often has rather similar results to induction, but it is far from instantaneous. For example, a gray target square whose background is changed from gray to red will immediately start to appear greenish. A few seconds later the new background will continue to appear red, but after two minutes or so this formerly red background will start to appear gray and the target square will appear greener than ever. During the first few seconds, I believe that the colour-appearance of the target square is primarily influenced by induction, and after tens of seconds to minutes, long-term chromatic adaptation increases in importance until it becomes the prime determinant of the target's appearance. Note that induction is asymmetric – the target's colour is influenced by the background, but the background's colour is unaffected. In contrast, long-term adaptation affects the colour of both background and target.

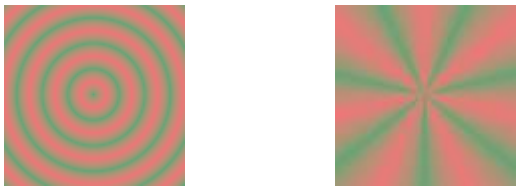
I believe that this distinction is particularly important, because the various methods of measuring colour-shifts probably invoke adaptation and induction to differing degrees. Commonly used techniques such as asymmetric matching or the choosing of reference-chips under free-viewing conditions are probably better suited to the study of adaptation than induction, and results determined using very different techniques such as nulling or short-duration 2AFC tasks are probably not directly comparable. It would be useful to have more information about the time-courses of adaptation and induction in order to design experiments that can address each phenomenon independently of the other.

Informal tests have given me reason to think that induction is affected by texture more strongly than adaptation is, and I have attempted to design experiments to demonstrate that this is so. Unfortunately this has been harder than I anticipated because of the problem in adapting the 2AFC method to study the moving target of colour-perception during adaptation, and because of the difficulty of remembering reference colours for periods as long as two minutes. But these problems are not insurmountable, and with further work I believe this may prove an illuminating area of research.

Studying the effects of texture on simultaneous contrast induction

Looking back over my thesis, my biggest regret is that I did not spend more time exploring texture-texture interactions in a bid to discover why my polychromatic / textured stimuli were not subject to contrast induction, and perhaps to relate contrast-induction to other known mechanisms in the visual system. Despite my result showing that contrast induction involves low-level mechanisms that are tuned to the cardinal axes, there is considerable uncertainty over where in the brain simultaneous contrast is processed. We still do not know whether it is processed in the retina, LGN, primary visual cortex, by several visual areas working together – or even by several mechanisms working independently.

The fact that simultaneous contrast does act at such a low-level of the visual system raises hope that it may ultimately be studied at a neurophysiological level. Indeed, some researchers such as Wachtler et al. have already studied center-surround interactions in Macaque primary visual cortex, that appear to be similar to chromatic induction (Wachtler, Sejnowski et al. 2003). A natural next step might be to investigate whether this population of neurons is also subject to the texture-interactions I have described, and whether they are tuned as narrowly to the cardinal axes.



a) *b)*

Figure 63: Example high-contrast radial and concentric targets. Induction in similar, lower-contrast stimuli could help determine whether induction is regulated by chromatic boundaries in textures, or by low-level texture segmentation.

My current results seem to show that edge-detection mechanisms are not important in the regulation of simultaneous contrast effects, however they are not conclusive and I feel that this area merits further investigation. Figure 63 shows example targets based on the inducing backgrounds investigated by Zaidi, Yoshimi and Flannigan (Zaidi, Yoshimi et al. 1991; Zaidi, Yoshimi et al. 1992) who found that induction by the background with concentric rings – similar to image (a) – was highly dependent on the spatial-frequency of the concentric rings in the background, but that induction by radial patterns such as pattern (b) was uninfluenced by spatial frequency.

One way of interpreting these results, would be that induction works 'radially' and is weakened because it must cross chromatic borders in backgrounds consisting of concentric rings, but remains strong in radially patterned backgrounds. If this is so it would be informative to compare induction in target 1a with induction in target 1b. One drawback of these stimuli is that the spatial frequency in the radially patterned stimuli varies from center to periphery, so it is not possible to equate the radial pattern with the concentric ring pattern in terms of spatial frequency. It would be necessary to investigate whether induction in these stimuli is dependent on spatial frequency, to determine whether or not it is valid to compare results from the two stimuli.

Contrast-contrast interactions

One of the hypotheses I called on to explain my result that texture in the background restored simultaneous contrast effects even when the target was still textured, was that texture-texture contrast between surround and target might nullify the effect of the texture in the target and cause the display to be equivalent to an uniform target viewed against a textured background.

Chubb, Solomon and Sperling have shown that texture-texture interactions are known to be tuned to spatial frequency and orientation (Chubb, Sperling et al. 1989), so a natural test for this hypothesis would be to investigate whether induction between textured targets and surrounds that differ either in spatial frequency, or in the orientation of anisotropic textures. Neurons that are tuned to orientation and spatial frequency have been studied in the LGN, so it is not inconceivable that the phenomena of contrast induction in uniform surfaces and texture-texture induction are in some way related.

Polychromaticity

A particular point raised by my experiments, is that many surfaces are multi-coloured ('*polychromatic*', to some) and these may be perceived differently from the monochromatic objects that are typically used in vision research. These findings mesh with the growing awareness that small-scale features of objects, such as granularity, interreflections on a fine or large-scale, and the scattering of light within complex surfaces (such as the skin, marble or the surfaces of fruits) may provide additional information about these surfaces, or about the light-sources illuminating them. In other words, the assumption that colour vision works the same way on uniform surfaces (such as the coloured papers or patches on a computer screen that have hitherto been studied in the lab) and complex surfaces may not be valid.

Many of these cues are novel, and have not yet been extensively investigated by computational neuroscientists. Yet polychromatic surfaces are not a world away from the 'Mondrians' and multicoloured scenes that computational researchers concentrated on in the past. Their primary difference is that the arrangement and reflectances of the facets of a polychromatic surface are in a fixed spatial configuration and are more likely to be under the same illumination – which is not true for different coloured objects spread more widely within a scene. Although this is not addressed by my experiments, the contrasts between different-coloured facets within a polychromatic scene may provide a colour-constant signature that could be used to identify it – even if the mean perceived colour of the surface were to change. Furthermore, the polychromatic surfaces in my experiments were perceived in a qualitatively different fashion from the uniform, monochromatic surfaces. Work on the statistics of natural images and non-classical receptive fields suggests that it is preferable to study the brain using those stimuli that most closely resemble the world it is designed to see, rather than those stimuli such as uniform surfaces and Gabor patches that are easiest for us to generate and describe mathematically.

Whilst the results of experiments 2.1-2.2 were not definitive, it seems likely that the effect of texture on simultaneous contrast is not directly related to polychromatic nature of the stimuli, which I believed may reinforce their colour-appearance against external influences. But even if the notion of polychromaticity is irrelevant to the interpretation of my experiments, it is an attribute common to many everyday objects and as such seems worthy of more thorough investigation than it has received thus far.

My control from experiment 2.1.2 provides some evidence that we are able to judge the 'overall' hue of polychromatic objects. This is an interesting finding that I did not altogether expect. It seems feasible that we may be called to make comparisons between the mean hues of objects with similar polychromatic textures, such as two apples whose skins both contain red and green patches but which are not equally ripe. However, it is harder to think of instances when the ability to compare the mean hues of dissimilar textures might prove valuable. Perhaps one such application is in 'material constancy' – experience shows that some textures can appear quite different as we view them from near or afar or in different parts of the visual field, and an ability to identify a surface by its 'mean colour' could help stabilise its appearance under different viewing conditions. Such 'just-so' stories can be dangerously seductive and it is with trepidation that I voice them here. However some ecological hypotheses such as the link between frugivory and trichromacy have proved testable (Parraga, Troscianko et al. 2000; Sumner and Mollon 2000), and hyperspectral imaging systems that are gradually becoming more readily available will doubtless play a role in the future study of polychromatic surfaces and their perception.

How do we perceive polychromatic surfaces?

An equally interesting question relating to of polychromatic surfaces, is that of how known neural mechanisms are involved in their perception. Some computational theories of colour-constancy invoke edge-detection and filling-in processes, and there is some evidence that neurons able to perform these functions can be found in the cortex. But in most of the polychromatic surfaces used in my experiments it was possible to judge the colour either of the whole surface, or of its individual elements. This may be more, less or equally true for naturally occurring polychromatic surfaces, but the observation stands that polychromatic surfaces contain numerous chromatic boundaries. Do 'edge-detector' cells process the boundaries within surfaces and the boundaries between surfaces differently? If so, is their action regulated by basic properties of the stimulus such as contrast or spatial frequency – or are they guided by top-down processes as we switch our examination of a surface between global and local levels?

Can polychromaticity aid colour-constancy?

Despite my belief that polychromaticity does not inhibit contrast induction by reinforcing local perceptual processes at the expense of global perceptual processes (i.e. by providing enough information for us to unambiguously determine the colour of a surface without reference to global features such as the colour of the background or illumination frameworks), I believe that polychromaticity may yet prove important in the study of colour-constancy.

Polychromatic surfaces provide more information than uniform surfaces

In the introduction, I discussed Land's demonstration that the colour of a paper viewed in the 'void condition' is determined by the combination of its reflectance and the colour of the illuminant – but that the colour of the same paper viewed as part of a 'Mondrian' scene remains roughly constant despite changes in the spectrum of the illuminant.

My 'Mondrian' patches were originally designed as miniature homages to Land's stimuli. There were some important differences – for example the colours of my stimuli were defined in terms of cone-contrasts and not as surface reflectance functions – but potentially there is no reason why a miniature Mondrian stimulus need provide any less information about the illuminant than a large Mondrian stimulus. In theory we may be able use these same algorithms to attain colour-constancy for some polychromatic surfaces, even when they are viewed in the void condition.

Polychromatic surfaces and specular highlights

A specular highlight on a monochromatic surface provides enough information to define a chromatic axis along which the illuminant must lie. But figure 64 shows that this is not sufficient to define the illuminant unambiguously – computational examinations of the problem show that the visual system must examine at least two specular highlights on two separate monochromatic surfaces in order to determine the point in colourspace where the two axes intersect, and hence determine the chromaticity of the illuminant.

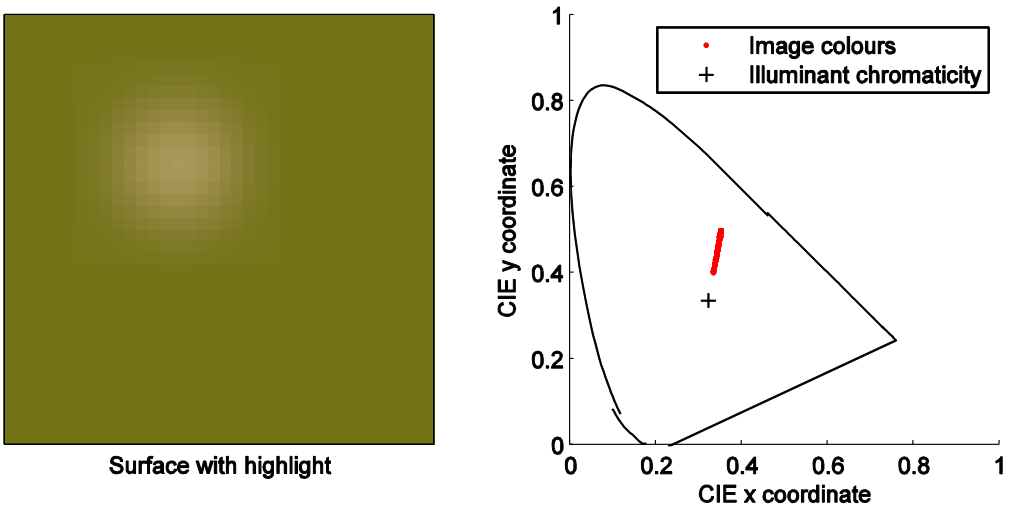


Figure 64: Specular highlights as a cue to illuminant chromaticity for an uniformly coloured surface

Left: An uniformly coloured surface, with a simulated specular highlight.

Right: The CIE 1931 chromaticity diagram, plotting the chromaticities present in the image to the left (red dots) and the chromaticity of the simulated illuminant (black cross). Note that both the red dots and the black cross are distributed along a common axis – but that it is impossible to determine where on this axis the illuminant should be plotted.

Clearly, a specular highlight on a polychromatic surface should provide better cues to the illuminant properties than a single specular highlight on a uniform surface – as is demonstrated in figure 64. But more importantly, a specular highlight on a polychromatic surface should provide more reliable cues than specular highlights on multiple, different-coloured uniform surfaces, as is demonstrated in figure 65. Some scenes contain multiple light sources, so if the visual system tries to combine information from multiple spatially separated specular highlights, it should ideally determine whether all of the specular highlights are from the same light source. For specular highlights on a polychromatic surface, this need not be necessary: purely local mechanisms could determine the chromaticity of the illumination falling on a surface, and correct the colour signal from the rest of the surface to achieve colour constancy.

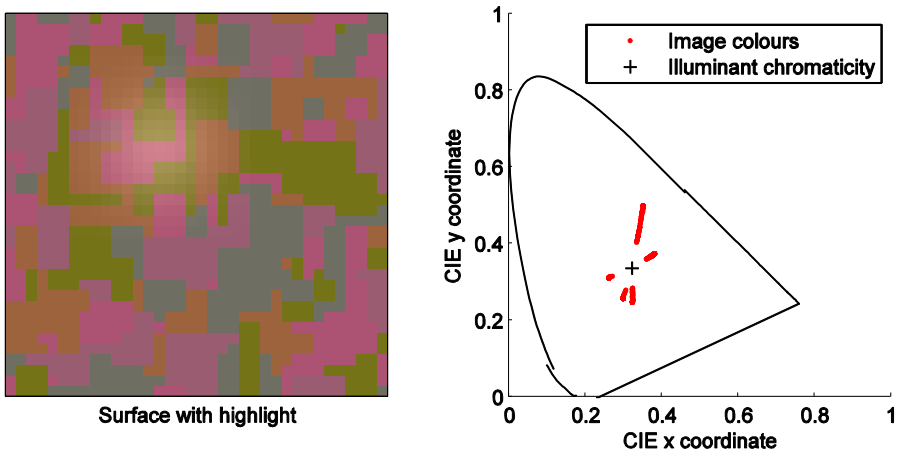


Figure 65: Specular highlights as a cue to illuminant chromaticity for an polychromatic surface

Left: A polychromatic surface, with a simulated specular highlight.

Right: The CIE 1931 chromaticity diagram, plotting the chromaticities present in the image to the left (red dots) and the chromaticity of the simulated illuminant (black cross). The polychromatic texture contains five basic colours, whose chromaticities can be plotted as five distinct points on the chromaticity diagram. The specular highlight draws these basic chromaticities towards a common center, that can be used to determine the chromaticity of the illuminant. In Illustration 2 it was impossible to determine the precise chromaticity of the illuminant, but by determining the common intersection of the five axes that may be easily distinguished on this graph, it is possible to unambiguously determine the chromaticity of the illuminant.

Given my enthusiasm for a role for polychromatic surfaces and specular highlights in attaining colour-constancy, I did generate some stereoscopic stimuli showing specular highlights against some small polychromatic surfaces, and I investigated them with the 2AFC technique. Preliminary results were disappointing and did not appear to show any major effects, so it may be that we do not have any faculty dedicated to the processing of highlights on polychromatic surfaces.

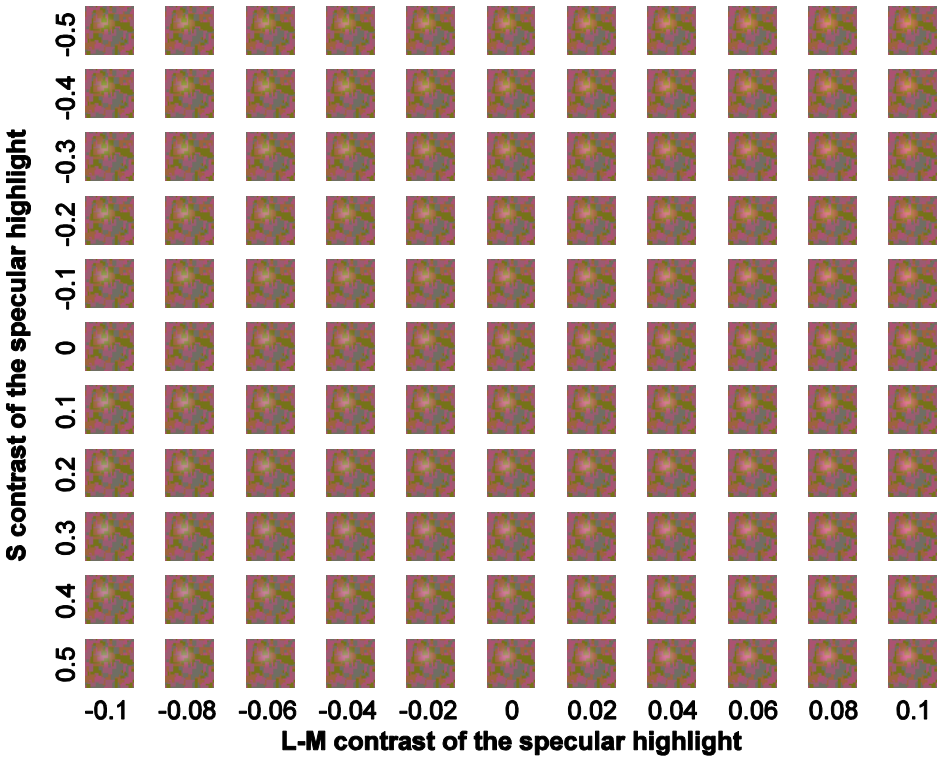


Figure 66: Identical polychromatic patches, with different-coloured specular highlights that should signify different illuminants

Figure 66 shows 121 identical polychromatic surfaces, each similar to the surface illustrated in figure 65 but boasting a different-coloured specular highlight. There is sufficient information in each patch to make a good estimate of the illuminant chromaticity, so if the local-processing hypothesis was correct, each of the surfaces should appear quite a different colour overall, but both 'eyeball' and preliminary experiments suggest that this is not the case.

This is hardly sufficient to prove that the visual system cannot extract information from specular highlights on polychromatic surfaces. Indeed researchers such as Maloney et al. have shown that specular highlights on uniformly coloured surfaces do influence our colour-perception (Yang and Maloney 2001). However further research is needed to determine whether the visual system does extract information from them, how it makes use of this information, and under what conditions it is able to do so.

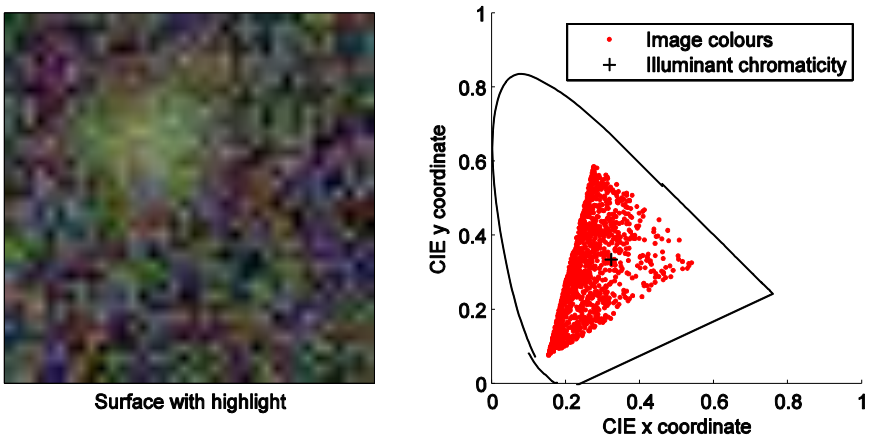


Figure 67: A specular highlight on a polychromatic surface with many component colours

It may be possible to estimate the chromaticity of the illumination from this highlight, but not by using the same strategy as shown in Illustration 4. This texture has many more component colours, so it is impossible to identify chromatic axes that intersect at the illuminant chromaticity.

As an example, figure 67 shows one situation where it may be difficult to extract information from specular highlights on polychromatic surfaces: namely when that polychromatic surface contains too many distinct colours. Perhaps the visual system can only extract information from surfaces that contain limited numbers of distinct colours. Another possibility is that many natural surfaces do not contain the correct distribution of component colours to provide reliable cues to the chromaticities of common illuminants. Again it would be interesting to examine polychromatic surfaces from a database of hyperspectral images to determine whether they can commonly provide useful information supporting colour-constancy. Even if the human visual system is unable to use such cues, they may prove useful in computer vision applications.

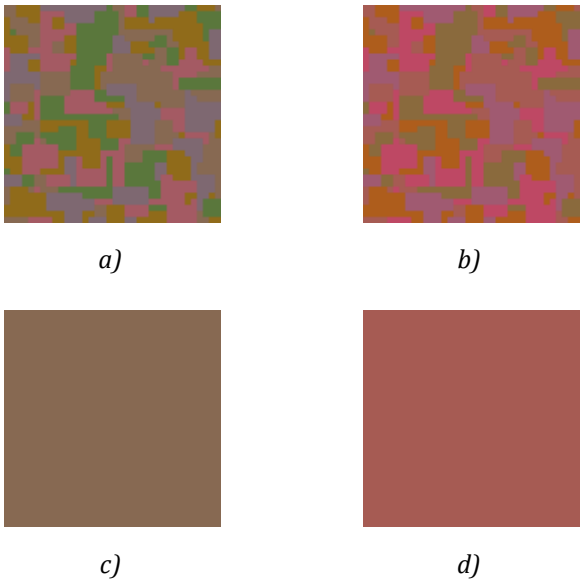


Figure 68: Polychromatic and uniform textures

Images in the same columns have the same mean chromaticities, and images in the same rows possess similar texture attributes, as well as the same cone-contrasts between the elements within individual images.

Images (a) and (b) show polychromatic surfaces with the same mean chromaticities as uniform surfaces (c) and (d) respectively. Yet, surfaces (a) and (b) appear more similar to each other than surfaces (c) and (d) – presumably due to the similar textures and the constant cone-contrasts present in the two polychromatic surfaces providing an additional clue to their identity.

A final observation on polychromaticity, is that it may provide surfaces with an extra chromatic signature that could be used to identify it or that could provide additional information about it, in addition to its 'mean chromaticity' – the property that colour-constancy tries to reinforce. Figure 68 illustrates how this might be the case. The basic prerequisite to colour constancy is that under changes in illumination, the surface chromaticities change in a predictable and consistent manner. Foster and Nascimento have shown that cone-contrasts between different surfaces are roughly constant under illumination changes, thus satisfying this condition (Foster and Nascimento 1994). Assuming that their finding also holds for the cone-contrasts within polychromatic surfaces, these internal contrasts may provide a signature that could be used to identify them despite failures in colour-constancy.

Appendix A – Observer Details

'Age' refers to the observer's age when they started to undertake observations for each experimental series.

The column 'naive' refers to the observer's knowledge of the experimental methods and aims. Observers marked 'Y' are naive; Observers marked 'N' are the experimenters and observers marked 'M; are moderately naive – i.e. they have some knowledge of psychophysics methods and/or colour theory. However, they are not working directly on the experiment and therefore do not have a detailed understanding of its objectives or methodology.

Experiment 1

Year of experiment: 2001

Table of observers

Initials	*Sex*	*FM score*	*FM pattern*	*Age*	*Naive*
KW	M	2	Normal	21	N
ACH	F	0	Normal	43	N
JK	F	16	Normal	20	Y
SW	F	12	Normal	22	Y
MH	M	8	Normal	23	Y

Table 34: Observer demographics and screening test scores for experiment 1

Additional notes:

KW and ACH were the experimenters. Observers JK, SW and MH were fine-art students who responded to an advertisement seeking paid observers.

Experiment 2

Year of experiment – 2002

Table of observers

Initials	*Sex*	*FM score*	*FM pattern*	*Age*	*Naive*
KW	M	2	Normal	22	N
ACH	F	0	Normal	44	N
YL	F	12	Normal	21	M
OC	M	?	Normal	?	Y
JW	M	22	Normal	21	M

Table 35: Observer demographics and screening test scores for experiment 2

Additional notes:

KW and ACH were the experimenters. Observer YL was a fellow PhD student in the lab. Whilst knowledgeable about colour science she did not have a detailed understanding of my project at the time she undertook the observations. Observers OC and JW were project students who undertook a set period of observing for me whilst working in the lab on separate projects. JW was working on projects related to colour-perception.

Experiment 3

Year of experiment: 2002-2004

Table of observers

INITIALS	*SEX*	*FM SCORE*	*FM PATTERN*	*AGE*	*NAIVE*
KW	M	2	Normal	22	N
ACH	F	0	Normal	45	N
YL (YZU)	F	12	Normal	22	M
AG	F	22	Normal	20	Y
AM	F	47	Normal	23	Y
LH	M	6	Normal	21	Y
LR	F	0	Normal	24	M
PM	M	14	Normal	22	Y
PMc	F	34	Normal	47	Y
PS	M	36	Normal	23	Y
SC	M	37	Normal	19	Y
JH	M	6	Normal	17	Y
JJN	M	8	Normal	21	M

Table 36: Observer demographics and screening test scores for experiment 3

Additional notes:

KW and ACH were the experimenters. Observer YL was a fellow PhD student in the lab. Whilst knowledgeable about colour science she did not have a detailed understanding of my project at the time she undertook the observations. Observers LR and JJN were students who were hired to run experiments in the lab. Observer PM worked as a secretary in the department. All other observers were paid students who responded to adverts placed around the university.

Appendix B – Choosing and Configuring Equipment for Use in Psychophysics Experiments, and a List of Equipment Used

Factors influencing the choice of computer displays

Until recently, cathode-ray-tube (CRT) based computer monitors were the only viable option for the generation of stimuli for vision-research. Several types of LCD monitors have been invented, but these are generally less well suited to the display of rapidly changing stimuli. Furthermore, the colours they display are very sensitive to changes in viewing angle, making such displays particularly unsuitable for colour-research. Data projectors offer the only other viable option - a number of different types have been invented based on CRT and LCD technologies and in recent years Texas Instruments "Digital Micromirror Devices" (DMD). Most can be used for vision research. However they are expensive both to buy and to run; LCD types can suffer stability problems and off-the-shelf DMD projectors are almost impossible to characterise (Seime and Hardeberg 2002; Wyble and Zhang 2003), so CRT monitors remain the default choice for most experimenters.

Trade secrets

The growing complexity of computer-systems may ultimately prove the biggest obstacle to accurate monitor calibration. A decade ago, it was possible for a single researcher to directly control a VGA graphics card at the hardware level, and the monitors available at that time were relatively simple affairs. Modern graphics cards have become vastly more capable than their forbears but their complexity has increased to match, and programmers are isolated from hardware by increasingly opaque levels of abstraction.

A 'feature' provided with many current NVidia graphics cards provides a typical example. When *Digital Vibrance*® is enabled, these cards use a proprietary algorithm that attempts to increase the apparent contrast and colour-saturation of their display (Nvidia 2002). Luckily, it can be turned off, and on the Linux operating system the feature is not currently enabled by default. NVidia cards have many other advanced functions. They are able to perform on-board gamma-correction or convert between floating point and integer colour representations, for example. It is entirely conceivable that the program used to calibrate a monitor may be run whilst the card is implementing gamma correction, but that the program used to display purportedly calibrated bitmap images is run with this feature disabled, or vice-versa. The result would be that the experiment was effectively run uncalibrated, and of course its findings may ultimately prove invalid.

Another example is that of Texas-Micromirror based data projectors. Calibration of these devices can be straightforward, and in many respects they have the potential to outperform other display technologies. But most small DLP projectors do not mix three primaries as CRT monitors and LCD projectors do. Rather, they use four – by mixing variable amounts of white with the coloured primaries, the maximum brightness the projector can produce is increased by 50%, and the results remains acceptable to all except psychophysicists, who can no longer use the simple additive calibration model (described in Chapter 3: Methods) that suffices for most other displays. The conversion between the computer's RGB colour space and the projector's RGBw colour-space is accomplished by circuitry inside the projector; there are generally no outward signs that a projector uses an RGBw colour-space. The inverse conversion cannot be performed without proprietary data that is not accessible to most experimenters (Seime and Hardeberg 2002; Wyble and Zhang 2003).

Most problems may probably be avoided by the simple expedient of checking a selection of the colours displayed in each experiment to ensure that they are close to those desired. But as the technical complexity of our tools increases exponentially and bugs and features become more insidious and more subtle, even the wary may be caught out on occasion. Psychophysicists have a strong tradition of technological subversion, and I strongly believe that it is in our best interests to use open technologies – programs to which we have the source code; peripherals whose design we thoroughly understand – whenever possible.

Equipment list and specifications

Computer system 1

Computer: Silicon Graphics Indy

Processor: 100 MHz MIPS

Memory: 64 Megabytes

Video resolution: 1280 x 1152

Operating system: Irix 4.x

Monitor

17" (non-Trinitron (i.e. with a curved screen))

Viewing apparatus

Viewing box with white diffusing interior; optionally fitted Wheatestone viewing apparatus.

Calibration apparatus

Minolta chromameter with RS232 interface box

Peripherals

Custom built button and dial interface with RS232 interface.

Software

Proprietary software written in C was used to perform both monitor characterisation and stimulus presentation.

Experiments

Used in experiment 1 only.

Computer system 2

Computer: Silicon Graphics O2

Processor: 180 MHz

Memory: 256 Megabytes

Video resolution: 1280 x 1152

Operating system: Irix 6.5x

Monitor

20” Silicon Graphics own brand (non-Trinitron)

Viewing apparatus

Viewing box with white diffusing interior; optionally fitted Wheatestone viewing apparatus.

Calibration apparatus

Minolta chromameter with RS232 interface box

Peripherals

Custom built button and dial interface with RS232.

Software

Proprietary software written in C was used to perform both monitor characterisation and stimulus presentation.

Experiments

Used in experiment 1 only.

Computer system 3

Computer: Desktop PC

Processor: 200 MHz Intel Pentium MMX

Memory: 196 Megabytes

Video card: Matrox Mystique

Video resolution: 800 x 600

Operating system: Windows 98

Monitor

20" Silicon Graphics own brand (non-Trinitron)

Software

Proprietary Matlab programs were used to perform display characterisation and to generate the stimuli in the form of bitmap images.

Software from Neurobehavioral systems (www.nbs.com) was used to perform stimulus presentation and response collection.

Experiments

Used in experiment 2.1

Computer system 4

Computer: Desktop PC (Intel)

Processor: 1.7 GHz Pentium 4

Memory: 256 Mb

Video card: NVidia GeForce 4 (64mb Video memory)

Video resolution: 800 x 600

Operating systems: Windows XP; Linux (Mandrake 8.2; 9.1; 9.2)

Monitor

HP 21" flatscreen (Trinitron) computer monitor

Viewing box: different for each experiment – see experimental notes for details

Software

Proprietary Matlab programs were used to perform display characterisation and to generate the stimuli in the form of bitmap images.

Software from Neurobehavioral systems (www.nbs.com) running under Windows was used to perform stimulus presentation and response collection in experiment 2. Custom-written Matlab software running under Linux was used to perform stimulus presentation, response collection and data-analysis in experiment 3.

Experiments

Experiments 2 & 3

Computer system 5

Computer: Desktop PC (Intel)

Processor: Intel Celeron 800 MHz

Memory: 256 Mb

Video card: NVidia GeForce 4 64Mb

Video resolution: 800 x 600

Operating systems: Linux (Mandrake 10)

Monitor

17" Ilyama Vision Master Pro 450 (flat-screen)

Viewing box: Viewing box with dark interior and Wheatstone stereoscope

Software

Proprietary Matlab programs were used to perform display characterisation and to generate the stimuli in the form of bitmap images.

Experiments

Experiment 3 (binocular disparity only)

Colorimetry equipment

Colorimeter

Minolta CS-100 (Minolta)

Experiments

Used in experiments 1-2

Spectrophotometer

Photo-Research PR-650 (Photo-Research 2009)

Experiments

Used in experiment 3

Appendix C - KCCV

Introduction

kccv ('Kit's Colour Conversions') is an extensible Matlab toolbox I wrote to convert colourspace co-ordinates between different formats. The functions it uses are mostly drawn from 'Measuring Colour' by Hunt (Hunt 2001) and from Wysecki & Style's book 'Color Science' (Wyszecki and Styles 1982). Unusual colour-space conversions are referenced in the code. The toolbox also includes a number of programs used to interface a Chromameter (Minolta CS100) or spectrophotometer (PR-650), and can use these to characterise monitors and test their resulting characterisation. Ancillary functions can be used to obtain information such as which calibration file or monitor is currently in use.

Design of KCCV

Matlab is optimised to work with matrixes, and has several other features that make *well* suited to processing colour-space co-ordinates. 'kccv' is designed to be both easy to use and efficient, and is able to convert matrixes of tristimulus co-ordinates between any two colourspaces with a single command.

The basic usage is:

>new_colour_matrix = kccv(colour_matrix, 'source_colourspace',…

'target_colourspace', ['whitepoint', whitepoint, …]);

The variable 'colour_matrix' may be a single triplet, an *n*x*3* vector (colourmap), an *n*x*m*x*3* image or even a four dimensional matrix of video frames. The only requirement is that the last dimension must have a length of 3.

Before processing the data, the 'kccv' wrapper function stores the original dimensions of the input file, then resizes it to create a long vector of tristimulus values. Internally standardising the data format makes it easier to write additional functions to extend the capabilities of the program. After performing the colour-space conversions, kccv returns the vector to its original dimensions.

To find a path by which arbitrary colour conversions can be performed, kccv generally uses the CIE 1931 colourspace (XYZ, Yxy) as an intermediary, and to add support for a new colourspace it is simply necessary to write two routines converting in both directions between XYZ and the new colourspace. It is also possible to specify alternative 'direct' routes for conversions where this is not necessary, in order to increase efficiency. Any new conversion method is written in the form of an 'm-file' with a filename composed of the source and target colourspaces. Detailed instructions on how to do this are incorporated into the source code of 'kccv.m'.

Performance

Matlab is an interpreted language and therefore performs many functions more slowly than compiled languages. However, if care is taken to avoid unnecessary loops and to allocate memory correctly, it can be nearly as efficient as a compiled language (Mathworks 2006). Kccv was written carefully and with efficiency in mind.

Although all the methods kccv employs are all well established and it does not constitute 'original research', it has proved to be a convenient tool that has greatly increased productivity in our laboratory.

The code: a kccv file index

I. Calibration routines
- Cal_init.m
- Calibrate.m
- Calltest.m
- Man_gamma.m

II. Main conversion routines
- kccv.m
- kmakergb.m
- kunmakergb.m

III. Conversion subroutines
- CCtoXYZ.m
- DBLRGBtoUI8RGB.m
- DKLtoXYZ.m
- LabtoCD.m
- LabtoLHC.m
- LabtoXYZ.m
- LHCtoLab.m
- LHCtoXYZ.m
- LHStoLUVPRIME.m
- LHStoXYZ.m
- LuvtoCD.m
- LUVPRIMEtoLHS.m
- LUVPRIMEtoXYZ.m
- LUVSTARtoLHS.m
- LUVSTARtoXYZ.m
- MCBtoXYZ.m
- tcc.m
- UI8RGBtoDBLRGB.m
- XYZtoCC.m
- XYZtoLab.m
- XYZtoLHC.m
- XYZtoLHS.m
- XYZtoLUVPRIME.m
- XYZtoLUVSTAR.m
- XYZtoMCB.m
- XYZtoYxy.m
- YxytoXYZ.m

IV. I/O and other files
- high_freq_test.m
- my_ginput.m
- pixelsize.m

Calibration routines

cal_init.m

```
1  function cal_init(varargin)
2  % Initialises all calibration variables - e.g. lookup tables, colourspace conversion matrices...
3  %
4  % ----------------------------------------------------------------------------------------
5  %
6  %  (C) 12/12/2001 Kit Wolf, dept of Physiology, University of Newcastle upon Tyne
7  %  Contact: c.j.l.wolf_AT_newcastle.ac.uk
8  %
9  %
10 %  Modifications/refinements: 08/02/2002, may be called with arguments 'GammaRampDataFile', name and
11 %            'MonitorDataFile', name so that the same computer may be used to
12 %            produce bitmaps with different calibrations for different monitors. KW
13 %
14 %            03/02/2004, creates a further dither_lut array to allow dithering for
15 %            find chromatic and brightness control
16 % ----------------------------------------------------------------------------------------
17
18
19
20 % Global variables section - these will be initialised, and made available to other functions
21 %
22 % By using an expanded lookup table it is possible to find the closest gun_value to the one we require
23 % with a minimum of computational effort
24 global gun_lut;
25 global dither_lut
26 global ideal_increment
27 % Raw calibration data, produced by program 'calibrate'
28 global CalData;
29 global GammaRampDataFile;
30 global MonitorDataFile;
31 global MonitorDat
32 % For removal of the 'baseline glow' - see later
33 global baseline;
34 % Matrices for the conversion of vectors in one colour space to vectors in another
35 global RGB_XYZ;
36 global XYZ_RGB;
37 global XYZ_LMS;
38 global LMS_XYZ;
39 %
40 % ----------------------------------------------------------------------------------------
41 % Load data from files
42 %
43 % default values,
44 GammaRampDataFile = 'ScreenCalDat';
45 MonitorDataFile = 'MonitorDat';
46 for i=1:nargin,
47   if strcmp(varargin(i), 'GammaRampDataFile'), GammaRampDataFile = varargin{i+1}; end
48   if strcmp(varargin(i), 'MonitorDataFile'), MonitorDataFile = varargin{i+1}; end
49 end
50 %
51 % Load calibration data generated by program 'Calibrate.m' - holds variable called CalData
52 load(GammaRampDataFile); % Holds raw data needed for gamma ramp correction
53 % Monitor data (screen size etc...) now incorporated into ScreenCalDat file - this line retained for
54 % backwards compatability
55 if ~isstruct(MonitorDat), load(MonitorDataFile); end
56 %
57 % ----------------------------------------------------------------------------------------
58 % Prepare variables for use in the expanded lookup table
59 %
60 % Allocate memory for large luts - by having several entries for each phosphor we can create an LUT
61 % that allows us to find the gun value that outputs the luminance closest to the one that we specified
62 % with a minimum of computational effort
```

```
63  gun_lut = uint8(zeros(3, 2048)); % 3: RGB; 5000 = entries in expanded lut - 20 per gun value is ample
64  % A separate 'dither_lut' remains in floating point format - that has a performance penalty but allows
65  % more accurate control over chromaticity and brightness.
66  dither_lut = zeros(3, 2048);
67  % Ideal_increment stores ideal luminance increment per step in the linearised lut
68  % - i.e. max.luminance/number_of_entries_in_LUT
69  ideal_increment = zeros(3, 1);
70
71  % ------------------------------------------------------------------------------------------------
72  % Gamma ramp lut interpolation
73  %
74  % A previous version of this program measured every gamma ramp value from 0 to 255 - later versions of
75  % 'calibrate' measure at intervals, rather than at every point. This means that calibration can be
76  % quicker, and hopefully more accurate due to reduced monitor drift whilst the measurements are being
77  % carried out. To retain compatibility, this interpolation routine moves CalData.readings so the names
78  % do not change.
79
80  if size(CalData.readings, 2) == 4, % only need to do all this if calibration file is new-style
81    CalData.measurements = CalData.readings;
82    CalData.readings = zeros(256, 3, 3);
83    for prim=1:3,
84      for Yxyparam=1:3,
85        CalData.readings(:, Yxyparam, prim) = interp1(CalData.measurements(:, 4, prim),
              CalData.measurements(:, Yxyparam, prim), 0:255, 'spline');
86      end
87    end
88  clear prim
89  clear Yxyparam
90  end
91
92  % ------------------------------------------------------------------------------------------------
93  % Baseline glow correction
94  %
95  % Correct the calibration data by removing baseline 'glow'
96  % 'baseline' holds XYZ coords of light 'glow' that is superimposed on each reading. Adding the 'glow' several
97  % times messes up the calibration, so we add it just once, when converting
98  baseline = kccv(mean(reshape(CalData.readings(1, :, :), 3, 3)'), 'Yxy', 'XYZ');
99  for i=256:-1:1, CalData.readings(i, 1, :) = CalData.readings(i, 1, :) - CalData.readings(1, 1, :); end
100 %
101
102 % ------------------------------------------------------------------------------------------------
103 % gun LUTs
104 %
105 % Prepare the expanded lookup tables. There are quicker, but more involved methods to do this
106 %
107 for gun=1:3
108         ideal_increment(gun) = (CalData.readings(256, 1, gun)/2047);
109         for i=1:2048
110                 [throwaway, closest_val] = min(abs(CalData.readings(:, 1, gun) – ...
((I1)*ideal_increment(gun))));
111                 gun_lut(gun, i) = closest_val-1;
112         end;
113 end
114
115 % ------------------------------------------------------------------------------------------------
116 % Dithering LUTs
117 %
118 % This section may ultimately replace the previous two, if dither_lut is replaced by gun_lut
119 %
120 %REMOVE THIS SECTION FOR NOW< AS IT DOES NOT WORK WITH CERTAIN CALDATA
    FILES (WITH REPEATED VALUES)
121 CalData.dither = CalData.measurements;
122
123 % Remove the baseline glow from these data
124 %for gun = 1:3,
```

```
125  %  CalData.dither(:, 1, gun) = CalData.dither(:, 1, gun) - CalData.dither(1, 1, gun);
126  %end
127
128  % Plot an endpoint - it helps the interpolation
129  %lastind = length(CalData.dither) + 1;
130  %for gun=1:3,
131  %  CalData.dither(lastind, :, gun) = [CalData.readings(256, :, gun), 255];
132  %end
133
134  % Create an interpolated lut_dither table
135  %for gun=1:3,
136  %  dither_lut(gun, :) = interp1(CalData.dither(:, 1, gun), CalData.dither(:, 4, gun), ...
(0:2047)*ideal_increment(gun), 'spline');
137  %end
138
139  %clear lastind
140  clear prim
141  clear Yxyparam
142  end
143
144
145  % ------------------------------------------------------------------------------------------------
146  % Create conversion matrices
147  %
148  % create the RGB to XYZ conversion matrix - in terms of luminance units
149  RGB_XYZ = ones(3, 3); % allocate temporary memory
150  for (i=1:3)       RGB_XYZ(i, 2) = CalData.readings(255, 2, i);
151                    RGB_XYZ(i, 3) = CalData.readings(255, 3, i);
152                    % convert to XYZ coordinates
153     RGB_XYZ(i, :) = kccv(RGB_XYZ(i, :), 'Yxy', 'XYZ');
154                    end
155  % input Rlum, Glum, Blum (array(1*3)); output XYZ coords of displayed colour (1*3)
156  %
157  % create matrix to convert colour specified in terms of XYZ primaries back to RGB
158  RGB_XYZ = RGB_XYZ'; % whoops
159  XYZ_RGB = inv(RGB_XYZ);
160  %
161  % Smith & Pokorny cone conversion matrix
162  XYZ_LMS = [0.15514       0.54312    -0.03286;
163    -0.15514       0.45684    0.03286;
164    0                             0                          0.01608];
165  %[0.45684, -0.15514, 0.03286; 0.15514, 0.54312, -0.03286; 0, 0, 0.01608];
166  LMS_XYZ = inv(XYZ_LMS);
167  %
168  % ------------------------------------------------------------------------------------------------
169  % leave file, and all global variables, in memory...
170  mlock;
171  % end of function
```

calibrate.m

```
1  function varargout = calibrate(varargin)
2  % CALIBRATE Application M-file for calibrate.fig
3  %  FIG = CALIBRATE launch calibrate GUI.
4  %  CALIBRATE('callback_name', ...) invoke the named callback.
5
6  % Last Modified by GUIDE v2.0 26-Mar-2002 13:27:48
7
8  if nargin == 0 % LAUNCH GUI
9
10         fig = openfig(mfilename,'reuse');
11   global CalData
12   load ScreenCalDat
13         % Use system color scheme for figure:
14         set(fig,'Color',get(0,'defaultUicontrolBackgroundColor'));
15
16         % Generate a structure of handles to pass to callbacks, and store it.
17         handles = guihandles(fig);
18   guidata(fig, handles);
19
20   if nargout > 0
21                   varargout{1} = fig;
22         end
23
24 elseif ischar(varargin{1}) % INVOKE NAMED SUBFUNCTION OR CALLBACK
25
26         try
27                   if (nargout)
28                             [varargout{1:nargout}] = feval(varargin{:}); % FEVAL switchyard
29                   else
30                             feval(varargin{:}); % FEVAL switchyard
31                   end
32         catch
33                   disp(lasterr);
34         end
35
36 end
37
38
39 %| ABOUT CALLBACKS:
40 %| GUIDE automatically appends subfunction prototypes to this file, and
41 %| sets objects' callback properties to call them through the FEVAL
42 %| switchyard above. This comment describes that mechanism.
43 %|
44 %| Each callback subfunction declaration has the following form:
45 %| <SUBFUNCTION_NAME>(H, EVENTDATA, HANDLES, VARARGIN)
46 %|
47 %| The subfunction name is composed using the object's Tag and the
48 %| callback type separated by '_', e.g. 'slider2_Callback',
49 %| 'figure1_CloseRequestFcn', 'axis1_ButtondownFcn'.
50 %|
51 %| H is the callback object's handle (obtained using GCBO).
52 %|
53 %| EVENTDATA is empty, but reserved for future use.
54 %|
55 %| HANDLES is a structure containing handles of components in GUI using
56 %| tags as fieldnames, e.g. handles.figure1, handles.slider2. This
57 %| structure is created at GUI startup using GUIHANDLES and stored in
58 %| the figure's application data using GUIDATA. A copy of the structure
59 %| is passed to each callback. You can store additional information in
60 %| this structure at GUI startup, and you can change the structure
61 %| during callbacks. Call guidata(h, handles) after changing your
62 %| copy to replace the stored original so that subsequent callbacks see
63 %| the updates. Type "help guihandles" and "help guidata" for more
```

```
64  %| information.
65  %|
66  %| VARARGIN contains any extra arguments you have passed to the
67  %| callback. Specify the extra arguments by editing the callback
68  %| property in the inspector. By default, GUIDE sets the property to:
69  %| <MFILENAME>('<SUBFUNCTION_NAME>', gcbo, [], guidata(gcbo))
70  %| Add any extra arguments after the last argument, before the final
71  %| closing parenthesis.
72
73
74
75  % --------------------------------------------------------------------
76  function varargout = Xdeg_edit_Callback(h, eventdata, handles, varargin)
77
78
79
80
81  % --------------------------------------------------------------------
82  function varargout = Ydeg_edit_Callback(h, eventdata, handles, varargin)
83
84
85
86
87  % --------------------------------------------------------------------
88  function varargout = Xpix_edit_Callback(h, eventdata, handles, varargin)
89
90
91
92
93  % --------------------------------------------------------------------
94  function varargout = Ypix_edit_Callback(h, eventdata, handles, varargin)
95
96
97
98
99  % --------------------------------------------------------------------
100  function varargout = stepsize_slider_Callback(h, eventdata, handles, varargin)
101  set(handles.stepsize_text, 'string', ['Number of measurements to take: ', ...
     num2str(round(get(handles.stepsize_slider, 'value')*255)+1)]);
102
103
104
105  % --------------------------------------------------------------------
106  function varargout = stepsize_text_Callback(h, eventdata, handles, varargin)
107
108
109
110
111  % --------------------------------------------------------------------
112  function varargout = go_button_Callback(h, eventdata, handles, varargin)
113
114  %  Do the calibration routine
115  CalData.operator = get(handles.experimenter_edit, 'string');
116  CalData.monitor = get(handles.monitor_edit, 'string');
117  CalData.computer = get(handles.computer_edit, 'string');
118  CalData.date = datestr(now, 0);
119  CalData.extra_notes = get(handles.notes_edit, 'string');
120  disp(sprintf('\nCalibration performed by %s on a %s monitor driven by %s on %s\n', ...
     CalData.operator, CalData.monitor, CalData.computer, CalData.date))
121
122  %  Draw a full-screen chromameter stimulus consisting of RGB (200, 200, 200) background
123  cal_display = figure;
124  set(cal_display, 'MenuBar', 'none', 'doublebuffer', 'on');
125  ScreenDims = get(0, 'Screensize');
126  set(cal_display, 'Position', [1, 1, ScreenDims(3:4)]);
```

```
scr = uint8(ones(12, 16, 3)*200);
set(gca, 'Position', [0 0 1 1]);
%  Allocate memory for chromameter measurements
%  256 phosphor levels, 4 readings (Y, little_x, little_y, R/G/B level, 3 guns) - cannot need more mem. than this
CalData.readings = zeros(256, 4, 3);
cps = cal_params;
CalSpectrum = zeros(256, 3, cps.num_spectral_points, 2);
%  Draw center patch (phosphor, level) to be measured & take readings
scr(5:7, 7:10, 1:3) = 0;
image(scr);
%  Wait a while before starting to take readings so that the monitor has time to stabilise
pause(60*(get(handles.calibration_start_slider, 'Value')*127));
current_reading_no = 0;
for level=0:256/round(get(handles.stepsize_slider, 'Value')*255+1):255
%  Interpolation seems to work well between readings, but extrapolates badly
  level = floor(level);
  current_reading_no = current_reading_no+1;
  for phosphor=1:3
    scr(5:7, 7:10, 1:3) = uint8(0);
    scr(5:7, 7:10, phosphor) = uint8(level);
    image(scr);
    figure(cal_display);
    set(gca, 'visible', 'off');
    text(2, 11, ['Current level: ', num2str(level)]);
    %  Allow time for image to be updated & to stabilise...
    pause(1);
    last_reading = cal_read;
    last_reading
    CalData.readings(current_reading_no, 1, phosphor) = last_reading.bigY
    CalData.readings(current_reading_no, 2, phosphor) = last_reading.x;
    CalData.readings(current_reading_no, 3, phosphor) = last_reading.y;
    CalData.readings(current_reading_no, 4, phosphor) = level;
    last_reading = cal_spectrum;
    if ~(strcmp(last_reading.errcaption, 'null')), CalSpectrum(current_reading_no, phosphor, :, :) ...
 = last_reading.spectrum; end
  end;
end;
CalData.readings = CalData.readings(1:current_reading_no, :, :);
CalSpectrum = CalSpectrum(1:current_reading_no, :,:,:);

%  Work out all the necessary values for 'MonitorDat' fields
MonitorDat.width = str2num(get(handles.Xdeg_edit, 'string'));   % width of picture on monitor screen
MonitorDat.height = str2num(get(handles.Ydeg_edit, 'string'));   % height of picture on monitor screen
MonitorDat.distance = str2num(get(handles.view_dist_edit, 'string'));
        % viewing distance from eye to monitor
MonitorDat.xpixels = ScreenDims(3); % pixels across monitor screen (horizontal)
MonitorDat.ypixels = ScreenDims(4); % pixels across monitor screen (vertical)
MonitorDat.xpd = MonitorDat.xpixels/(atan(MonitorDat.width/MonitorDat.distance)/(2*pi)*360);
% number of pixels reqd. to display 1 degree horizontal line
 MonitorDat.ypd = MonitorDat.ypixels/(atan(MonitorDat.height/MonitorDat.distance)/(2*pi)*360); % number of pixels reqd. to display 1 degree vertical line
MonitorDat.pd = mean([MonitorDat.xpd; MonitorDat.ypd]);
% number of pixels required to display 1 degree long line - avg of x/y dimensions

%  Save and clear up...
save(get(handles.filename_edit, 'string'), 'CalData', 'MonitorDat');
if ~(strcmp(last_reading.errcaption, 'null')), save([get(handles.filename_edit, 'string'), 'Spectrum'], ...
        'CalSpectrum', 'CalData'); end
close(cal_display);
disp(sprintf('\nCalibration complete\n'))

%  Plot gamma ramps of previously acquired calibration data
load(get(handles.filename_edit, 'string'));
axes(handles.red_gamma);
```

```
184  plot(CalData.readings(:,4,1), CalData.readings(:,1,1), 'r.-')
185  legend('Red primary', 2)
186  legend boxoff
187  axes(handles.green_gamma);
188  plot(CalData.readings(:,4,2), CalData.readings(:,1,2), 'g.-')
189  legend('Green primary', 2)
190  legend boxoff
191  axes(handles.blue_gamma);
192  plot(CalData.readings(:,4,3), CalData.readings(:,1,3), 'b.-')
193  legend('Blue primary', 2)
194  legend boxoff
195  load([get(handles.filename_edit, 'string'), 'Spectrum']);
196
197  %  Plot the spectral data
198  current_reading_no = size(CalSpectrum, 1);
199  axes(handles.red_spectral);
200  plot(reshape(CalSpectrum(current_reading_no, 1, :, 1), 1, size(CalSpectrum, 3)), ...
reshape(CalSpectrum(current_reading_no, 1, :, 2), 1, size(CalSpectrum, 3)), 'r')
201  legend('Red primary', 0)
202  legend boxoff
203  axes(handles.green_spectral);
204  plot(reshape(CalSpectrum(current_reading_no, 2, :, 1), 1, size(CalSpectrum, 3)), ...
            reshape(CalSpectrum(current_reading_no, 2, :, 2), 1, size(CalSpectrum, 3)), 'g')
205  legend('Green primary', 0)
206  legend boxoff
207  axes(handles.blue_spectral);
208  plot(reshape(CalSpectrum(current_reading_no, 3, :, 1), 1, size(CalSpectrum, 3)), ...
            reshape(CalSpectrum(current_reading_no, 3, :, 2), 1, size(CalSpectrum, 3)), 'b')
209  legend('Blue primary', 0)
210  legend boxoff
211
212  %  Change any of the text strings (e.g. operator, monitor, computer...)
213  set(handles.monitor_edit, 'string', CalData.monitor);
214  set(handles.experimenter_edit, 'string', CalData.operator);
215  set(handles.computer_edit, 'string', CalData.computer);
216
217  %  Write details into the edit box
218  edit{1} = ['Date: ', CalData.date];
219  edit(Kraft and Brainard) = ['Red primary coords: x=' num2str(CalData.readings(current_reading_no, 2, 1)), ' y=',
...
            num2str(CalData.readings(current_reading_no, 3, 1))];
220   edit(Kraft and Brainard) = ['Green primary coords: x=' num2str(CalData.readings(current_reading_no, 2, 2)), '
y=', ...
            num2str(CalData.readings(current_reading_no, 3, 2))];
221  edit{4} = ['Blue primary coords: x=' num2str(CalData.readings(current_reading_no, 2, 3)), ' y=', ...
            num2str(CalData.readings(current_reading_no, 3, 3))];
222  edit(H. Munsell) = ['Red primary max. brightness: Y=' num2str(CalData.readings(current_reading_no, 1, 1)), ' ...
            Cd/M^2'];
223  edit(Kraft and Brainard) = ['Green primary max. brightness: Y=' num2str(CalData.readings(current_reading_no,
1, 2)), ' ...
            Cd/M^2'];
224  edit(Kraft and Brainard) = ['Blue primary max. brightness: Y=' num2str(CalData.readings(current_reading_no, 1,
3)), ' ...
            Cd/M^2'];
225  edit{8} = CalData.extra_notes
226  set(handles.further_info, 'string', edit);
227
228
229  % --------------------------------------------------------------------
230  function varargout = filename_edit_Callback(h, eventdata, handles, varargin)
231
232
233
234
235  % --------------------------------------------------------------------
```

```
236 function varargout = experimenter_edit_Callback(h, eventdata, handles, varargin)
237
238
239
240
241 % --------------------------------------------------------------------
242 function varargout = computer_edit_Callback(h, eventdata, handles, varargin)
243
244
245
246
247 % --------------------------------------------------------------------
248 function varargout = monitor_edit_Callback(h, eventdata, handles, varargin)
249
250
251
252
253 % --------------------------------------------------------------------
254 function varargout = notes_edit_Callback(h, eventdata, handles, varargin)
255
256
257 % --------------------------------------------------------------------
258 function varargout = slider3_Callback(h, eventdata, handles, varargin)
259 set(handles.calibration_start_delay_text, 'string', ['Start delay: ', ...
        num2str(round(get(handles.calibration_start_slider, 'Value')*127)), ' minutes']);
260
261 % --------------------------------------------------------------------
262 function varargout = load_button_Callback(h, eventdata, handles, varargin)
263
264 %  Plot gamma ramps of previously acquired calibration data
265 load(get(handles.filename_edit, 'string'));
266 axes(handles.red_gamma);
267 plot(CalData.readings(:,4,1), CalData.readings(:,1,1), 'r.-')
268 legend('Red primary', 2)
269 legend boxoff
270 axes(handles.green_gamma);
271 plot(CalData.readings(:,4,2), CalData.readings(:,1,2), 'g.-')
272 legend('Green primary', 2)
273 legend boxoff
274 axes(handles.blue_gamma);
275 plot(CalData.readings(:,4,3), CalData.readings(:,1,3), 'b.-')
276 legend('Blue primary', 2)
277 legend boxoff
278 load([get(handles.filename_edit, 'string'), 'Spectrum']);
279
280 %  Plot the spectral data
281 current_reading_no = size(CalSpectrum, 1);
282 axes(handles.red_spectral);
283 plot(reshape(CalSpectrum(current_reading_no, 1, :, 1), 1, size(CalSpectrum, 3)), ...
        reshape(CalSpectrum(current_reading_no, 1, :, 2), 1, size(CalSpectrum, 3)), 'r')
284 legend('Red primary', 0)
285 legend boxoff
286 axes(handles.green_spectral);
287 plot(reshape(CalSpectrum(current_reading_no, 2, :, 1), 1, size(CalSpectrum, 3)), ...
        reshape(CalSpectrum(current_reading_no, 2, :, 2), 1, size(CalSpectrum, 3)), 'g')
288 legend('Green primary', 0)
289 legend boxoff
290 axes(handles.blue_spectral);
291 plot(reshape(CalSpectrum(current_reading_no, 3, :, 1), 1, size(CalSpectrum, 3)), ...
        reshape(CalSpectrum(current_reading_no, 3, :, 2), 1, size(CalSpectrum, 3)), 'b')
292 legend('Blue primary', 0)
293 legend boxoff
294
295 %  Change any of the text strings (e.g. operator, monitor, computer...)
296 set(handles.monitor_edit, 'string', CalData.monitor);
```

```
297  set(handles.experimenter_edit, 'string', CalData.operator);
298  set(handles.computer_edit, 'string', CalData.computer);
299
300  %  Write details into the edit box
301  edit{1} = ['Date: ', CalData.date];
302  edit(Kraft and Brainard) = ['Red primary coords: x=' num2str(CalData.readings(255, 2, 1)), ' y=', ...
            num2str(CalData.readings(255, 3, 1))];
303  edit(Kraft and Brainard) = ['Green primary coords: x=' num2str(CalData.readings(255, 2, 2)), ' y=', ...
            num2str(CalData.readings(255, 3, 2))];
304  edit{4} = ['Blue primary coords: x=' num2str(CalData.readings(255, 2, 3)), ' y=', ...
            num2str(CalData.readings(255, 3, 3))];
305  edit(H. Munsell) = ['Red primary max. brightness: Y=' num2str(CalData.readings(255, 1, 1)), ' Cd/M^2'];
306   edit(Kraft and Brainard) = ['Green primary max. brightness: Y=' num2str(CalData.readings(255, 1, 2)), '
Cd/M^2'];
307  edit(Kraft and Brainard) = ['Blue primary max. brightness: Y=' num2str(CalData.readings(255, 1, 3)), ' Cd/M^2'];
308  edit{8} = CalData.extra_notes
309  set(handles.further_info, 'string', edit);
310
311
312  % --------------------------------------------------------------------
313  function varargout = view_dist_edit_Callback(h, eventdata, handles, varargin)
314  %if get(str2num(handles.view_dist_edit, 'string'),
```

caltest.m

```
1   %  Does calibration test, using screen calibration functions
2   %  09/10/01 KW, University of Newcastle upon Tyne
3   %
4   %  Possible bug: ensure that the computer running the calibration uses RGB (24 / 32 bit)
5   %  colour - on windows look at 'display properties.' On my colormap SGI the default is to
6   %  raster the display(!ouch)
7   %
8   %  Requires matlab v. 6 (R 12) or higher to work (serial support). Will not yet work under Irix.
9   %
10  global RGB_XYZ
11  global CalData
12
13  %  Start by obtaining info about calibrator
14  TstData.operator = input('Who is doing this caltest?\n', 's');
15  TstData.date = datestr(now, 0);
16  disp(sprintf('\nConfirm correct: caltest performed by %s on a %s monitor driven by %s\n Y/N', ...
            CalData.operator, CalData.monitor, CalData.computer))
17  if ~strcmp(upper(input('', 's')), 'Y') error('Sorry...'); end
18  %
19  %  prepare matrix of randomly chosen screen coordinates
20  testlength = input('How many measurements do you want?');
21  TstData.testcols = uint8(rand(1, testlength, 3)*255);
22  TstData.predicted = zeros(1, testlength, 3);
23  TstData.measured = zeros(1, testlength, 3);
24  %
25  %  convert them into predicted colours
26  %for i=1:3, TstData.predicted(1, :, i) = CalData.readings(double(TstData.testcols(1, :, i))+1, 1, i); end
27  TstData.predicted = kccv(TstData.testcols, 'UI8RGB', 'Yxy');
28  TstData.predicted
29
30  disp('Pausing for 10 seconds...');
31  pause(10);
32
33  %
34  %  ---------------------------------------------------------------------------
35  %
36  %  Display and measure 'em
37  %
38  %  ---------------------------------------------------------------------------
```

```
%
%  draw a full-screen chromameter stimulus consisting of RGB (200, 200, 200) background
figure
set(gcf, 'MenuBar', 'none');
set(gcf, 'Position', get(0, 'screensize'));
scr = uint8(ones(12, 16, 3)*200);
set(gca, 'Position', [0 0 1 1]);
%  allocate memory for chromameter measurements
%  draw center patch (phosphor, level) to be measured & take readings
for measurement=1:length(TstData.testcols),
   for i=1:3, scr(5:7, 7:10, i) = TstData.testcols(1, measurement, i); end
   image(scr);
   set(gca, 'visible', 'off');
   %  allow time for image to be drawn & to stabilise...
   pause(1);
   %  take reading
   last_reading = cal_read;
   TstData.measured(1, measurement, 1) = last_reading.bigY;
   TstData.measured(1, measurement, 2) = last_reading.x;
   TstData.measured(1, measurement, 3) = last_reading.y;
  end;
%  save and clear up...
save('CalTstOutput', 'TstData');
close(gcf);
disp(sprintf('\nCaltest complete\n'))
```

man_gamma.m

```
function equiv_value = man_gamma(varargin)
% MAN_GAMMA Application M-file for man_gamma.fig
%  FIG = MAN_GAMMA launch man_gamma GUI.
%  MAN_GAMMA('callback_name', ...) invoke the named callback.

% Last Modified by GUIDE v2.0 24-May-2002 16:31:41
global which_primary_to_configure;
global equiv_value;
if nargin ==1 & isnumeric(varargin{1}), if varargin{1} < 4, which_primary_to_configure = varargin{1}; ...
else which_primary_to_configure = 1; end
        fig = openfig(mfilename,'reuse');
        % Use system color scheme for figure:
        set(fig,'Color',get(0,'defaultUicontrolBackgroundColor'));

        % Generate a structure of handles to pass to callbacks, and store it.
        handles = guihandles(fig);
        guidata(fig, handles);

        % Wait for callbacks to run and window to be dismissed:
        uiwait(fig);
%       if nargout > 0
%               varargout{1} = fig;
%       end

elseif ischar(varargin{1}) % INVOKE NAMED SUBFUNCTION OR CALLBACK

        try
                if (nargout)
                        [varargout{1:nargout}] = feval(varargin{:}); % FEVAL switchyard
                else
                        feval(varargin{:}); % FEVAL switchyard
                end
        catch
                disp(lasterr);
        end

```

```
36 end
37
38
39 %| ABOUT CALLBACKS:
40 %| GUIDE automatically appends subfunction prototypes to this file, and
41 %| sets objects' callback properties to call them through the FEVAL
42 %| switchyard above. This comment describes that mechanism.
43 %|
44 %| Each callback subfunction declaration has the following form:
45 %| <SUBFUNCTION_NAME>(H, EVENTDATA, HANDLES, VARARGIN)
46 %|
47 %| The subfunction name is composed using the object's Tag and the
48 %| callback type separated by '_', e.g. 'slider2_Callback',
49 %| 'figure1_CloseRequestFcn', 'axis1_ButtondownFcn'.
50 %|
51 %| H is the callback object's handle (obtained using GCBO).
52 %|
53 %| EVENTDATA is empty, but reserved for future use.
54 %|
55 %| HANDLES is a structure containing handles of components in GUI using
56 %| tags as fieldnames, e.g. handles.figure1, handles.slider2. This
57 %| structure is created at GUI startup using GUIHANDLES and stored in
58 %| the figure's application data using GUIDATA. A copy of the structure
59 %| is passed to each callback. You can store additional information in
60 %| this structure at GUI startup, and you can change the structure
61 %| during callbacks. Call guidata(h, handles) after changing your
62 %| copy to replace the stored original so that subsequent callbacks see
63 %| the updates. Type "help guihandles" and "help guidata" for more
64 %| information.
65 %|
66 %| VARARGIN contains any extra arguments you have passed to the
67 %| callback. Specify the extra arguments by editing the callback
68 %| property in the inspector. By default, GUIDE sets the property to:
69 %| <MFILENAME>('<SUBFUNCTION_NAME>', gcbo, [], guidata(gcbo))
70 %| Add any extra arguments after the last argument, before the final
71 %| closing parenthesis.
72
73
74
75 % --------------------------------------------------------------------
76 function varargout = slider1_Callback(h, eventdata, handles, varargin)
77 draw_comparison_graphs(handles);
78
79
80
81 % --------------------------------------------------------------------
82 function varargout = slider2_Callback(h, eventdata, handles, varargin)
83 draw_comparison_graphs(handles);
84
85
86
87 % --------------------------------------------------------------------
88 function varargout = Done_button_Callback(h, eventdata, handles, varargin)
89 close(gcf)
90
91
92 function draw_comparison_graphs(handles)
93
94 % take reading from the sliders
95 global which_primary_to_configure;
96 global equiv_value;
97
98 equiv_value = round((get(handles.slider1, 'Value')*245 + get(handles.slider2, 'Value')*10));
99
100  axes(handles.axes1);
```

```
101  testim = zeros(1,1,3);
102  testim(:,:,which_primary_to_configure) = equiv_value/255;
103  image(testim);
104  set(handles.axes1, 'visible', 'off');
105
106  % work out screen height of lhs axes (rhs axes should be equal)
107  set(handles.axes1, 'Units', 'Pixels');
108  size = get(handles.axes1, 'Position');
109  height = size(4);
110  width = size(3);
111  clear size;
112
113  testim = zeros(height, width, 3);
114  testim(1:2:height,:,which_primary_to_configure) = 1;
115
116  % now draw everything
117  axes(handles.axes2);
118  image(testim);
119  axis image;
120  set(handles.axes2, 'visible', 'off');
121
122  % and return the data
```

Main conversion routines

kccv.m

```
1  % --------------------------------------------------------------------------------------------------------------------------------------------
2  % KCCV - A FREE MATLAB TOOLBOX FOR COLOURSPACE CONVERSIONS AND COMPUTER MONITOR AND PRINTER CALIBRATION
3  % --------------------------------------------------------------------------------------------------------------------------------------------
4  %
5  % USAGE - A QUICK GUIDE
6  % If the data is held in a matrix, array or vector...
7  % Basic usage is 'dataout = kccv(mydata, 'present_format', 'target_format', 'whitepoint', [Y, x, y], ...
       otheroptions...);
8  % The 'whitepoint' need only be specified for certain conversions, e.g. cone_contrast or Luv
9  %
10 % If the data matrix is held in a structure...
11 % The data matrix (with the same format as for the simple function) is held in the field 'data'
12 % The present data format is held in a field with the name 'type' (e.g. mystruct.type = 'RGB')
13 %
14 % The functions are generally very efficient in converting many tristimulus vectors (e.g. arrays; matrixes) at once. The functions are grossly inefficient at
15 % converting single tristimulus vectors one after another - this is matlab, after all.
16 %
17 % KCCV LICENSING AGREEMENT
18 % This function, 'kccv' converts points, vectors or arrays of values between different colourspaces. It was originally written by Kit Wolf, with contributions
19 % from Yazhu Ling
20 %
21 % USING THE TOOLBOX
22 % 'kccv' acts as a gateway to a library of subfunctions that operate only on n*3 vectors of tristimulus values. It may take inputs in two formats -
23 % either as Matlab matrices (1*3, n*3, n*m*3... dimensional), or as structures - which contain a field named 'data' which holds a normal input matrix, and 'type'
24 % which holds a string specifying the current format of the data in 'data.' The toolbox is intended primarily for neuroscientists studying colour - but is designed
25 % to be modular, so that additional functionality may easily be incorporated into the program. If you make substantial modifications which you feel may be useful
26 %
27 %
28 %
29 % SUPPORTED FORMATS
30 % Yxy - In terms of CIE 1931 colourspace coordinates
31 % XYZ - In terms of the powers of the standard X, Y and Z illuminants (CIE, 1931)
32 % LMS - In terms of the excitations of the L, M and S cones ('red', 'green' and 'blue' cones) - currently calculated using the Smith-Pokorny conversion matrices
33 % RGB - device dependent RGB
34 % UI8RGB - device dependent RGB - in 24 bit truecolour format
35 % DBLRGB - matlab colour format - as double precision tristimulus values between 0 and 1 - calculated using
36 % PUI8RGB - calibrated colour for inkjet printers (and possibly other printers)
37 % CC - cone-contrast space; A whitepoint must be specified
38 % MCB - McCloed-Boyton-like colourspace; A whitepoint must be specified
39 % Lab - A whitepoint must be specified
40 % Luvprime - A whitepoint must be specified
41 % Luvstar - A whitepoint must be specified
42 % LHS - A whitepoint must be specified
43 % LHC - A whitepoint must be specified
44 % CD - colour difference formulae for Luv and Lab colour spaces. Functions - LuvtoCD and LabtoCD
45 %
46 % For more information on conversions between each format, type 'help FORMAT' at the command line. e.g. 'help MCB'
47 %
48 % RGB FORMATS AND CALIBRATION
49 % 'kccv' can convert data between absolute colorimetric and device-dependent formats: for example between
```

```
Yxy and XYZ and 24 bit RGB. For this to be accurate,
% you must first calibrate your output device (colour CRT monitor or data projector) accurately. The toolbox incorporates a number of tools that can help with
% this - notably a spectrophotometer interface and the function 'calibrate.' We also incorporate a monitor calibration interface that does not require use of
% potentially expensive calibration instruments. This allows reasonably accurate calibration for less-demanding applications.
%
% Gamma correction is achieved using an expanded lookup table with 2048 entries. This means that all 16 million available output RGB values are potentially
% available.
%
% CALIBRATED COLOUR IN MATLAB, WINDOWS and X11
% The accuracy of this program is dependent on the assumption that the output program (for example, a program written in DirectX which loads and displays
% ready-made bitmaps) and the program used to calibrate the monitor - typically 'calibrate,' provided with this library, and run under matlab - do not modify
% or correct the gamma ramp differently from each other.
%
% COLOUR INDEXING AND IMAGE FORMATS
% Windows and X11 may use dithering to try and display more realistic images, if running in a colorindex mode. SGI 'irix' opens indexed-colour windows by
% default. Truecolor modes should be enabled, and dithering disabled on such systems. Care should also be taken when saving images in file formats other than
% non-lossy 24 bit truecolor.
%
% BUGS
% This version contains only undiscovered bugs. Although Yazhu Ling and myself have been using the toolbox for some time, I have just undertaken a major rewrite
% and it is possible that some bugs have been introduced at this stage. If you discover any new ones, please let me know by emailing c.j.l.wolf@ncl.ac.uk
%
% AND 'FEATURES'...
% The 'Digital Vibrance' setting on NVIDIA graphics cards (and potentially equivalents on other brands) may substitute the colours you choose for ones that appear
% more pleasing to the eye. Luckily it can be turned off. As graphics cards and packages such as openGL become more complex, they are provided with ever more
% settings that could invalidate the calibration provided by these functions.
%
%
% KCCV FOR HACKERS
% 'KCCV' is the interface to a range of functions, and in most cases, additional functionality may be added without modifying this file. We ask that you do not
% do so, as this may mean other users cannot use functions that you add. Even if you do not plan to make your functions available to others, modifying kccv.m
% may make your functions are incompatible with later releases of 'KCCV'. If you write code that absolutely requires modifications to 'KCCV,' then send it to us,
% and we will attempt to incorporate it into later releases of the program.
%
% Most tristimulus conversions covered by this program are achieved by transforming tristimulus vectors to and from XYZ tristimulus vectors. This allows a
% simple interface to be designed whereby the majority of colourspace conversions (e.g. Yxy to RGB) may be achieved through two transformations using XYZ values
% as an intermediate. In some instances, this cannot be achieved - for example the conversion from XYZ to LHS involves an intermediate conversion through Luv.
% The program gets round this through an additional function, 'XYZtoLHS', which calls the functions 'XYZtoLuv' and 'LuvtoLHS' in turn. Finally, in some instances
% it may be desirable to use conversions that do not use XYZ as an intermediate at all. For example, in comparing two sets of LHS values, it is clearly more
% efficient to bypass the conversion through Luv to XYZ, or it may be that we wish to improve the efficiency of a function by gluing two matrix conversions back
% to back and eliminating an useless conversion through the XYZ intermediate. For allow for these cases, the program first looks for a conversion function named
% [fromTRP, 'to', 'toTRP] - if it does not find one, it will call a function named [fromTRP, 'toXYZ'], then one named ['XYZto', toTRP]. By entering a conversion
```

```
91  % as the first argument after the input data, matrix conversions may be performed directly and the program
          will set the label of the returned data type as the
92  % second conversion label (toTRP).
93  %
94  % Finally, all functions pass all input arguments to kccv to each colour conversion function. In this way, it is
          possible to add extra functionality to individual
95  % functions without changing the functioning of the remainder of the program, ensuring backwards
          compatibility accross releases.
96  %
97  function im = kccv(im, varargin)
98  prevwarningstate = warning('off');
99  % ------------------------------------------------------------------------------------------------------------------------
100  % section 1 - determine the current & target formats of the data in 'im.' Convert 'im' to nxmx3 matrix ---------
101  % ------------------------------------------------------------------------------------------------------------------------
102  %
103  % interpret the command line
104  if isstruct(im),  fromTRP = im.type;
105          use_structure_format = 1;
106          if nargin<2, error('Function ccv requires target triplet type to be specified'), end
107          toTRP = varargin{1};
108        else
109          if nargin<3, error('Function ccv requires at least 3 input arguments - if ''im'' is not a structure ...
- im, current_format & target_format'), end
110          fromTRP = varargin{1};
111          toTRP = varargin(Kraft and Brainard);
112          use_structure_format = 0;
113          im.data = im;
114        end
115  %
116  % check that init_cal in memory; if it isnt, then run it
117  if isempty(strmatch('cal_init', inmem)), cal_init; end
118  %
119  % make global variables from 'cal_init'
120  global XYZ_LMS
121  global LMS_XYZ
122  global XYZ_RGB
123  global RGB_XYZ
124  %
125  % remember the original size of 'im'
126  orig_size = size(im.data);
127  % reshape 'inim' from 1*3 or n*m*3... to an n*3 matrix, to make 'kccv' an universal function -
128  % able to take single, vector, and array sets of colourspace coordinates as its arguments
129  if orig_size(length(orig_size))~=3, error('The last dimension of the input matrix must have a length of ...
''3'''), end
130  im.data = reshape(im.data, prod(orig_size)/3, 3);
131  %
132  % ------------------------------------------------------------------------------------------
133  % section 2 - convert the data in 'im' to another format ------------------------------
134  % ------------------------------------------------------------------------------------------
135  %
136  % If a direct-access function exists, then use it:
137  if ~isnumeric(fromTRP) & ~isnumeric(toTRP) & exist([fromTRP, 'to', toTRP]), im.data = ...
feval([fromTRP, 'to', toTRP], im.data, varargin);
138   %
139   % otherwise, first convert the image into XYZ matrix - the 'go-between' format
140   else
141   if ~isnumeric(fromTRP),
142   switch lower(fromTRP),
143     case('lms'), im.data = im.data*LMS_XYZ';
144     case('rgb'), im.data = im.data*RGB_XYZ';
145     case('yxy'), im.data = YxytoXYZ(im.data);
146     case('ui8rgb'), im.data = KunmakeRGB(im.data); % for 24 bit colour modes
147     case('dblrgb'), im.data = KunmakeRGB(im.data*255);
% for 24 bit colour modes - assume that the RGB data was scaled to lie between 0 and 1
148     case('xyz'), ; % do nothing
```

```
149   otherwise im.data = feval([upper(fromTRP), 'toXYZ'], im.data, varargin);
% allows extra functionality to be added with no need to edit this file
150   end
151   else im.data = im.data*fromTRP; end
152   %
153   % finally convert the image into whatever target format has been specified
154   if ~isnumeric(toTRP),
155   switch lower(toTRP),
156     case('lms'), im.data = im.data*XYZ_LMS';
157     case('rgb'), im.data = im.data*XYZ_RGB';
158     case('yxy'), im.data = XYZtoYxy(im.data);
159     case('ui8rgb'), im.data = KmakeRGB(im.data, varargin, 'orig_imsize', orig_size);
% useful for 24 bit colour modes
160     case('dblrgb'), im.data = KmakeRGB(im.data, 'double', varargin, 'orig_imsize', orig_size);
% matlab demands that figure colormaps be arrays of double precision integers between 0 and 1
161     case('xyz'), ; % do nothing
162   otherwise im.data = feval(['XYZto', upper(toTRP), ], im.data, varargin);
163   end
164   else im.data = im.data*toTRP; end
165  end
166  %
167  % ------------------------------------------------------------------------------------------
168  % section 3 - convert the data in 'im' to the correct output format ------------------
169  % ------------------------------------------------------------------------------------------
170  %
171  %  Make output matrix the same size as the input matrix
172  im.data = reshape(im.data, orig_size);
173  %
174  % if the input was not a structure, then use the structure with the same fields
175  im.type = toTRP;
176  if ~use_structure_format, im = im.data; end
177  %
178  % reset the warning state as it is a global rather than local variable
179  warning(prevwarningstate);
180  % end of function. koniec. fin. vzeo
```

KmakeRGB.m

```
1   function img = KmakeRGB(img, varargin)
2   % takes matrix of double, XYZ values; 'weeds' out undisplayable pixels and
3   % uses lookup tables from 'init_cal' to do gamma correction.
4   % returns uint8 (24 bit color) matrix
5   %
6   %
7   % The following variables were set in 'cal_init'
8   global dither_lut
9   global gun_lut
10  global ideal_increment
11  global baseline
12  global XYZ_RGB
13  % The persistent variable 'spat_lum' is used to correct for spatial inhomogeneities
14  persistent spat_lum
15  %
16  % remove the XYZ component of the image that will be supplied by the baseline glow
17  for i=1:3, img(:, i) = img(:,i) - baseline(i); end
18  %
19  % convert the image into RGB primaries
20  img = kccv(img, 'XYZ', 'RGB');
21  % ADDED LINE TO ENSURE FILENAME COMPATIBILITY
22    [mynum, myind] = max(strcmp(varargin, 'orig_imsize'));
23    orig_imsize = size(img); % This holds the original dimensions of the img array (in this function)
24    temp_imsize = varargin{myind+1};
% And this holds the dimensions of the img array passed as an argument to kccv (i.e. 2d array of pixels)
25  %
```

```
% ----------------------- Unique feature - correction for spatial non-uniformities ---------------------------------
% check whether we want to correct for spatial non-uniformities? (is spat_correct specified)
[mynum, myind] = max(strcmp(varargin{1}, 'spat_correct'));
if mynum == 1,
  % check that all parameters req'd for spatial-correction have been specified correctly
  if (length(varargin{1}) < myind+2 & isnumeric(varargin{1}{myind+1}) &
isnumeric(varargin{1}{myind+2})),
    error('Usage: kccv(usual_input_arguments, ''spat_correct'', xoffset, yoffset)')
  end
  % if so, then go!
  disp('Implementing spatial corrections');
  if isempty(spat_lum), load spat_corrections.mat; end
  xoffset = varargin{1}{myind+1};
  yoffset = varargin{1}{myind+2};
  % alas, img needs to be reshaped to implement spatial correction
  [mynum, myind] = max(strcmp(varargin, 'orig_imsize'));
  orig_imsize = size(img); % This holds the original dimensions of the img array (in this function)
  temp_imsize = varargin{myind+1};
% And this holds the dimensions of the img array passed as an argument to kccv (i.e. 2d array of pixels)
  img = reshape(img, temp_imsize);
  for i=1:3,
    img(:,:,i) = img(:,:,i).*spat_lum(xoffset:xoffset+size(img, 1)-1, yoffset:yoffset+size(img, 2)-1, i);
  end
  img = reshape(img, orig_imsize);
  %  if xoffset + size(img, 1) > size(spat_lum, 1), error('Image cropped to fit screen'
end
clear mynum myind
%
% ------------------------ All done (well, almost) ------------------------------------------------------------------
%
for i=1:3, img(:, i) = img(:, i)./ideal_increment(i); end
%
if (max(strcmp('RGB_Ceil', varargin)) == 1),
  % remove the undisplayable and replace with either 0 or ceiling values
  img(isnan(img)) = 0;
  img(img(:,:) > 2047) = 2047;
  img(img(:,:) < 0) = 0;
else
  % prepare a mask recording the locations of those values that are within the gamut
  weedmask = img(:,1)<2047 & img(:,2)<2047 & img(:,3)<2047 &img(:,1)>=0 & ...
img(:,2)>=0 & img(:,3)>=0;
  % set these to '0'
  for i=1:3 img(:,i) = img(:,i).*weedmask; end
end
%
% round the entries so they can serve as indexes to a matrix
img = round(img);
% remove nan values (little x or y values == 0);
img(isnan(img)) = 0;
%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%% Dithering section %%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Check whether we wish to 'dither' the image to improve the chromatic and luminance resolution
[mynum, myind] = max(strcmp(varargin{1}, 'dither'));
if mynum == 1,
% This piece of code alternates lines of two different colours. Uint8 'floors' any value passed to it.
% So example value 20.25 is rounded down to '20' - as is value 20.5, 20.99 etc...
% Alternating lines with value 20 with value 21 gives a mean colour of 20.5 - so the range
% 19.75-20.25 should be rendered exclusively as value 20
% 20.25-20.75 should be rendered as value 20.5 (alternated 21 / 20)
% 20.75-21.25 should be rendered as value 21
%
% Adding 0.25 & 0.75 to alternate lines should achieve this.
```

```
88  %
89    myargs = varargin{1};
90    randomise = 0;
91    if length(myargs)>myind, if strcmp(lower(myargs{myind+1}), 'random'), randomise = 1; else ...
randomise = 0; end; end
92    if randomise == 0,
93      dithermask = ones(size(img)).*0.25;
94      dithermask(1:2:size(img, 1), :) = 0.75;
95    else
96  % For complete randomisation, a value of 20 should be rendered exclusively as 20;
97  % For a value of 20.1, 9/10 pixels should be rendered as '20' and 1/10 rendered as 21
98  % Adding a random number between 0 and 1 should achieve this.
99      dithermask = rand(size(img));
100 %  else?!
101 % Other forms of dithering are possible, and may be desirable in some instances. For example,
102 % complete randomisation may give rise to perceptible chromatic textures on a small scale.
The use of a regular
103 % array that allows more than an extra bit of resolution is possible; the higher the chromatic resolution, the
104 % larger the spatial grid must be.The use of regular patterns may cause artifacts on some stimuli - such as
105 % thin lines.
106   end
107   for i=1:3, img(:,i) = dither_lut(i, reshape(img(:, i), size(img, 1), 1)+1)'; end
108   img = img + dithermask;
109 else
110 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
111 %%%%%%% Non-dithering section %%%%%%
112 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
113   for i=1:3, img(:,i) = gun_lut(i, reshape(img(:, i), size(img, 1), 1)+1)'; end
114 end
115 %
116 % remove the undisplayable and replace with [0,0,0] based on the gamut mask we prepared earlier
        (weedmask)
117 % due to possible noise in the low-luminance readings, the lowest values prepared
118 if (max(strcmp('RGB_Ceil', varargin)) == 0),
119  for i=1:3 img(:,i) = img(:,i).*weedmask; end
120 end
121
122 if nargin==1, varargin{1} = 'uint8'; end
123 %if strcmp(varargin(Kraft and Brainard), 'double'), img = floor(img)/255; else img = uint8(img-1); end %%ACH's
change
124 if strcmp(varargin(Kraft and Brainard), 'double'), img = img/255; else img = uint8(img); end
125
126 % ---------------------------- Why is the spatial re-calibration only almost finished? -------------------------------
127 % Calls to 'image' display bitmaps with the 1st dimension on the Y-axis and 2nd dimension on X-axis – i.e.
128 % rotated 90 degrees. Likewise when we use 'imwrite' - an 800 x 600 image ends up in portrait mode. As
        most spatially
129 % calibrated bitmap images are ultimately for display on a computer monitor and to avoid the potential for
        flipping / rotating
130 % the images the wrong way by hand, I prefer to resize the images here and save them within the function.
131 [mynum, myind] = max(strcmp(varargin{1}, 'filename'));
132 if mynum == 1,
133   if ischar(varargin{1}{myind+1}), filename = varargin{1}{myind+1}; else ...
error('A filename must be specified'); end
134   [mynum, myind] = max(strcmp(varargin{1}, 'filetype'));
135   if mynum ~=1, filetype = 'bmp'; else filetype = varargin{1}{myind+1}; end
136   tempimg = reshape(img, temp_imsize);
137   tempimg = imrotate(tempimg, 90);
138   imwrite(tempimg, [filename, '.', filetype], filetype);
139 end
140 % end of function
```

KunmakeRGB.m

```
1   function img = unmakeRGB(img)
```

```
2  % takes matrix of uint8, RGB values;
3  % uses lookup tables from 'init_cal' (CalData.readings) to convert to XYZ
4  % returns double matrix
5  %
6  global baseline
7  global RGB_XYZ
8  global CalData
9  %
10 % convert the image into type 'double'
11 img = double(img);
12 %
13 % convert the image into RGB luminances
14 for i=1:3, img(:,i) = CalData.readings(reshape(img(:,i), size(img(:,i), 1), 1)+1, 1, i); end
15 %
16 % convert into XYZ
17 img = kccv(img, 'RGB', 'XYZ');
18 %
19 % correct for baseline 'glow'
20 for i=1:3, img(:,i) = img(:,i) + baseline(i); end
21 % end of function
```

Conversion subroutines

CCtoXYZ.m

```
1  %  cmat is a n*3 matrix of L, M, and S-cone contrasts;
2  %  Whitepoint must be specified.
3  %  refcol = 1*1*3 vector of Yxy values
4  %  Lnew = (1+Lcontrast)*Lold; likewise for M, S contrasts
5  % Kit Wolf Jan 2002, Dept. of Physiology
6  function im = CCtoXYZ(cmat, varargin)
7
8  % prepare varargin, which may have got mangled in the passing
9  while iscell(varargin{1}), varargin = varargin{1}; end
10
11 % search for the input argument specifying the whitepoint
12 %if length(varargin)<2, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
13 for i=1:length(varargin)-1,
14   if ischar(varargin{i}),
15     if strcmp('whitepoint', lower(varargin{i})), refcol = kccv(varargin{i+1}, 'Yxy', 'LMS'); end
16   end
17 %  if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
18 end
19
20 im = zeros(size(cmat, 1), 3);
21 for i=1:3, im(:,i) = refcol(i); end
22 for i=1:3, im(:,i) = im(:,i) .* (1+cmat(:,i)); end
23 im = kccv(im, 'LMS', 'XYZ');
```

DBLRGBtoUI8RGB.m

```
1  % function dbl = DBLRGBtoUI8RGB(ui8, varargin)
2  % This function converts double precision RGB to UI8RGB, bypassing conversions to XYZ etc etc.
3  %
4  function ui8 = DBLRGBtoUI8RGB(dbl, varargin)
5
6  ui8 = uint8(dbl.*255);
```

deltaE.m

```
1  function deltaE=deltaE(colour1, space1, colour2, space2, whiteYxy)
2
3  % Obtain the Luv colour differences between colour1 and colour2
4  % Colour1 and colour2 must have the same size, eiter n*3 or n*m*3, which is accepted by kccv
5  % useage deltaE(colour1, space1, colour2, space2, whiteYxy)
6  % for example, deltaE([30 .21 .22], 'Yxy', [30.22 .24], 'Yxy', [50 .321 .337])
7  % Written by Yazhu Ling at 19/April/2004
8
9
10
11 si1=size(colour1);
12 si2=size(colour2);
13 if si1~=si2, error('colour1 and colour2 have to have the same size.'); end;
14 if si1(length(si1))~=3, error('The last dimension of the input matrix must have a length of 3.'); end;
15
16 Luv1=kccv(colour1, space1, 'luvstar', 'whitepoint', whiteYxy);
17 Luv2=kccv(colour2, space2, 'luvstar', 'whitepoint', whiteYxy);
18
19
20 if length(si1)==2
21   deltaE=sum((Luv1-Luv2).^2,2).^0.5;
22 elseif length(si1)>2
23   Luv1=reshape(Luv1, prod(si1)/3, 3);
24   Luv2=reshape(Luv2, prod(si2)/3, 3);
25   deltaE=sum((Luv1-Luv2).^2,2).^0.5;
```

```
26   si_deltaE=si1;
27   si_deltaE(length(si1))=1;
28   deltaE=reshape(deltaE, si_deltaE);
29  end;
```

DKLtoXYZ.m

```
1  %  cmat is a n*m*3 matrix of luminance, LM contrasts and S-cone contrasts; refcol = 1*1*3 vector of Yxy
          values
2  %  function im = contrast_convert(cmat, refcol)
3  function im = MCBtoXYZ(cmat, varargin)
4
5  error('This function has not yet been completed');
6
7  % prepare varargin, which may have got mangled in the passing
8  while iscell(varargin{1}), varargin = varargin{1}; end
9
10  % search for the input argument specifying the whitepoint
11  for i=1:length(varargin)-1,
12    if ischar(varargin{i}),
13      if strcmp('whitepoint', lower(varargin{i})), refcol = kccv(varargin{i+1}, 'Yxy', 'XYZ'); end
14    end
15  %  if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
16  end
17  if ~exist('refcol', 'var'), error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
18
19  % fill the matrix with the LMS values of the whitepoint
20  im = zeros(size(cmat, 1), 3);
21  for i=1:3, im(:,i) = refcol(i); end
22  im = kccv(im, 'XYZ', 'LMS');
23
24  % brightness values
25  for i=1:3, im(:,i) = im(:,i).*(cmat(:,1)+1); end
26
27  % isoluminant L-M axis
28  for i=1:size(cmat, 1),
29     delta = tcc(im(i, 1), im(i,2), cmat(i, 2));
30     im(i, 1) = im(i, 1) + delta;
31     im(i, 2) = im(i, 2) - delta;
32  end
33
34  % S axis
35  im(:,3) = im(:,3).*(1+cmat(:,3));
36
37  % reconvert to XYZ
38  im = kccv(im, 'LMS', 'XYZ');
```

LabtoCD.m

```
1  % Function returns a measure of how great the perceptual difference between two Lab specified colours is.
2  % Function retunrns an n*3 matrix of values, the second two columns of which are set to '0'. This is to
3  % ensure that all reshapes are performed correctly.
4  % (C) Kit WOlf, University of Newcastle upon Tyne, 2002
5  %
6  function CD = LabtoCD(Lab, varargin);
7  varargin = varargin{1};
8  % search for the input argument specifying the whitepoint
9  if length(varargin)<2, error('A comparison set of colours must be specified. ...
Format: ''ComparisonSet'', matrixname'); end
10  for i=1:length(varargin)-1,
11    if ischar(varargin{i}),
12      if strcmp('comparisonset', lower(varargin{i})), CompSet = varargin{i+1}; end
13    end
```

```
14  %  if i==length(varargin)-1, error('A comparison set of colours must be specified. ...
Format: ''ComparisonSet'', matrixname'); end
15  end
16  CD = zeros(size(Lab));
17  CD(:,1) = ((CompSet(:,1)-Lab(:,1)).^2 + (CompSet(:,2)-Lab(:,2)).^2 + (CompSet(:,3)-Lab(:,3)).^2).^0.5;
18  % Formulae taken from "Measuring colour" by R.W.G. Hunt, 3rd edition, Fountain press, Kingston upon
          Thames, UK, 1998, pg. 275
19  % The end
```

LabtoLHC.m

```
1   % This function convert the CIELab values L, a and b into Luminance
2   % Hue and Chroma. Both Lab and LHC are n*3 matrix
3   % (C) Kit Wolf, University of Newcastle, 5/4/2002
4   function LHC=LabtoLHC(Lab);
5
6   LHC = zeros(size(Lab)); % allocate memory
7
8   % Do the 'L'
9   LHC(:, 1) = Lab(:, 1);
10
11  % Do the 'H'
12  LHC(:, 2) = atan2(Lab(:,3), Lab(:,2));
13  LHC(:, 2) = LHC(:, 2) + ((LHC(:,2)<0).*2.*pi);
 % This line changes range of atan2 to 0-2*pi rather than -pi-pi
14
15  % Do the 'C'
16  LHC(:, 3) = (Lab(:, 2).^2+Lab(:, 3).^2).^0.5;
17
18  % Koniec
```

LabtoXYZ.m

```
1   % This fucntion convert the CIELab colour space into tristimulus values
2   % The Lab is a n*3 matrix represent CIEL, a, b values
3   % The function return a 1*3 matrix XYZ
4   % XnYnZn is a 1*3 array representing the reference white in cie Yxy coordinates
5   % (C) 17/12/2001 by Yazhu Ling in University of Newcastle
6   % Converted for use with kccv in 21/06/2002 by Kit Wolf
7   function XYZ = LabtoXYZ(Lab, varargin);
8   varargin = varargin{1};
9   % search for the input argument specifying the whitepoint
10  if length(varargin)<2, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
11  for i=1:length(varargin)-1,
12    if ischar(varargin{i}),
13      if strcmp('whitepoint', lower(varargin{i})), XnYnZn = kccv(varargin{i+1}, 'Yxy', 'XYZ'); end
14    end
15  %  if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
16  end
17  %
18  sz = size(Lab);
19  conv_F = zeros(sz); % Allocate memory
20  conv_I = zeros(sz); % Allocate memory
21  XYZ = zeros(sz); % Allocate memory
22  clear sz
23
24  conv_F(:, 2) = (Lab(:, 1)+16)/116;
25  conv_F(:, 1) = conv_F(:, 2)+(Lab(:, 2)/500);
26  conv_F(:, 3) = (200*conv_F(:, 2)-Lab(:, 3))/200;
27
28  temp = conv_F<0.206893;
29  if conv_I(~temp) ~= [], conv_I(~temp) = conv_F.^3; end
30  if conv_I(temp) ~= [], conv_I(temp) = (conv_F-(16/116))./7.787; end
```

```
31
32  for i=1:3, XYZ(:, i) = conv_I(:, i)*XnYnZn(i); end
```

LHCtoLab.m

```
1   % This function converts the Luminance, Hue and Chroma into tristimulus value
2   % The L H C values are lightness, hue and chroma values in the CIELab colour space in an n*3 matrix
3   % The function return Lab values in an n*3 matrix
4   %
5   % 17/12/2001 by Yazhu Ling in University of Newcastle
6   % 05/04/2002 modifications by Kit Wolf, of University of Newcastle upon Tyne - to cope with n*3 arrays for
          use with kccv
7   function Lab = LHCtoLab(LHC);
8
9   Lab = LHC; % memory allocation
10
11  tmp_mat = LHC(:, 2)<0 | LHC(:, 2)>2*pi;
12  LHC(tmp_mat, 2) = 2*pi+rem(LHC(tmp_mat, 2),2*pi);
13
14  %if LHC(2)<0
15  %  LHC(2)=2*pi+rem(LHC(2),2*pi);
16  %end
17  %if LHC(2)>2*pi
18  %  LHC(2) = rem(LHC(2),2*pi);
19  %end
20  Lab(1) = LHC(1);
21
22  for i=1:size(LHC, 1),
23    %Lab(1) = LHC(1);
24    if LHC(i,2)==0.5*pi
25      Lab(i,2) = 0;
26      Lab(i,3) = LHC(i,3);
27    elseif LHC(i,2)==1.5*pi
28      Lab(i,2) = 0;
29      Lab(i,3) = -LHC(i,3);
30    elseif (LHC(i,2)<1.5*pi)&(LHC(i,2)>0.5*pi)
31      Lab(i,2) = -(((LHC(i,3)^2)/(1+tan(LHC(i,2))^2))^0.5);
32      Lab(i,3) = Lab(i,2)*tan(LHC(i,2));
33    else
34      Lab(i,2) = ((LHC(i,3)^2)/(1+tan(LHC(i,2))^2))^0.5;
35      Lab(i,3) = Lab(i,2)*tan(LHC(i,2));
36    end
37  end
```

LHCtoXYZ.m

```
1   % This function convert LHC to XYZ
2   % LHC represent Lightness, Hue and Chroma in CIELab colour space
3   % Both the input and the output are the 1*3 matrix
4
5   function cvar=LHCtoXYZ(cvar, varargin) % original Cvar = LHC, of course
6   varargin = varargin{1};
7   cvar=LHCtoLab(cvar);
8   cvar=LabtoXYZ(cvar, varargin);
9
```

LHStoLUVPRIME.m

```
% This function converts the Luminance, Hue and Saturation into tristimulus values
% The L H S values are lightness, hue and saturation values in the CIELuv colour space in an n*3 matrix
% The function return Luv values in an n*3 matrix
%
% 10/02/2003 by Kit Wolf, of University of Newcastle upon Tyne - to cope with n*3 arrays for use with kccv
% Formula derived from "Measuring colour" by R.W.G. Hunt, 3rd edn., Fountain Press, Kingston-upon-
Thames, England(1998). pg. 276
function Luv = LHStoLUVPRIME(LHS, varargin);

% prepare varargin, which has got mangled in the passing
if length(varargin) > 0, while iscell(varargin{1}), varargin = varargin{1}; end, end

% search for the input argument specifying the whitepoint
if ~strcmp(varargin, 'whitepoint'), error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
for i=1:length(varargin)-1,
  if ischar(varargin{i}),
    if strcmp('whitepoint', lower(varargin{i})), XnYnZn = prime_it(kccv(varargin{i+1}, 'Yxy', 'XYZ')); ...
          break; break; end
  end
  if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
end

Luv = LHS; % memory allocation

LHS(:,2) = mod(LHS(:,2), pi*2) + log(2)/10^14; % modulus divide the hue values to set their range 0 to 2*pi

Luv(:, 3) = 1; % assume that v is positive
Luv(0.5*pi <= LHS(:, 2) & LHS(:, 2) <= 1.5*pi, 3) = -1; % but if v should be negative, make it so

Luv(:,2) = tan(LHS(:, 2));
Luv(pi/2 <= LHS(:,2) & LHS(:,2) <= 1.5*pi, 2) = Luv(pi/2 <= LHS(:,2) & LHS(:,2) <= 1.5*pi, 2).*-1;


% using LHS(:,3) as a scratch variable to avoid having to allocate more memory
% work out the ratio of the actual length of the h/s vector to the length of the h/s vector we got when
LHS(:,3) = (LHS(:,3)./13)./sqrt((1 + Luv(:,2).^2));

Luv(:, 2) = Luv(:, 2).*LHS(:,3);
Luv(:, 3) = Luv(:, 3).*LHS(:,3);

% add u'n & v'n to (u' - u'n) and (v' - v'n) to get u' & v'
Luv(:, 2) = Luv(:, 2) + XnYnZn(2);
Luv(:, 3) = Luv(:, 3) + XnYnZn(3);

% and that's it.

function primed = prime_it(XtoZ)

denom = (XtoZ(:,1) + XtoZ(:,2)*15 + XtoZ(:,3)*3);
primed = zeros(size(XtoZ));
primed(:, 1) = XtoZ(:,2);
primed(:, 2) = (XtoZ(:,1)*4)./denom; % Calculate u' (or un')
primed(:, 3) = (XtoZ(:,2)*9)./denom; % Calculate v' (or vn')
```

LHStoXYZ.m

```
1  % This function convert Luv to XYZ
2  % Luv represent Lightness, Hue and Saturation in CIELuv colour space
3  % Both the input and the output are the 1*3 matrix
4
5  function cvar=LuvtoXYZ(cvar, varargin) % original Cvar = Luv, of course
6  %varargin = varargin{1};
7  cvar=LHStoLUVPRIME(cvar, varargin);
8  cvar=LUVPRIMEtoXYZ(cvar, varargin);
```

LUVPRIMEtoLHS.m

```
1  % This function convert the CIELuv values L, a and b into Luminance
2  % Hue and Saturation. Both Luv and LHS are n*3 matrix
3  % (C) Kit Wolf, University of Newcastle, 5/4/2002
4  % Formula from "Measuring colour"
by R.W.G. Hunt, 3rd edn., Fountain Press, Kingston-upon-Thames, England(1998). pg. 276
5  function LHS=LUVPRIMEtoLHS(Luv, varargin);
6
7  % prepare varargin, which has got mangled in the passing
8  if length(varargin) > 0, while iscell(varargin{1}), varargin = varargin{1}; end, end
9
10 % search for the input argument specifying the whitepoint
11 if ~strcmp(varargin, 'whitepoint'), error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
12 for i=1:length(varargin)-1,
13   if ischar(varargin{i}),
14     if strcmp('whitepoint', lower(varargin{i})), XnYnZn = prime_it(kccv(varargin{i+1}, 'Yxy', 'XYZ')); ...
           break; break; end
15   end
16   if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
17 end
18
19 LHS = Luv; % allocate memory
20
21 % Do the 'L'
22 %LHS(:, 1) = Luv(:, 1);
23
24 % Do the 'H'
25 LHS(:, 2) = atan2(Luv(:,2) - XnYnZn(2), Luv(:,3) - XnYnZn(3));
26 LHS(:, 2) = LHS(:, 2) + ((LHS(:,2)<0).*2.*pi);
          % This line changes the range of atan2 to 0-2*pi rather than -pi to pi
27
28 % Do the 'S'
29 LHS(:, 3) = 13*((Luv(:, 2) - XnYnZn(2)).^2 + (Luv(:, 3) - XnYnZn(3)).^2).^0.5;
30 %LHS(:, 3) = ((Luv(:, 2).^2+(Luv(:, 3).^2)).^0.5)./Luv(:,1);
31
32 % Koniec
33
34 function primed = prime_it(XtoZ)
35
36 denom = (XtoZ(:,1) + XtoZ(:,2)*15 + XtoZ(:,3)*3);
37 primed = zeros(size(XtoZ));
38 primed(:, 1) = XtoZ(:,2);
39 primed(:, 2) = (XtoZ(:,1)*4)./denom; % Calculate u' (or un')
40 primed(:, 3) = (XtoZ(:,2)*9)./denom; % Calculate v' (or vn')
```

LUVPRIMEtoXYZ.m

```
% This fucntion convert the CIEluv colour space into tristimulus values
% The luv is a n*3 matrix represent CIEL, a, b values
% The function return a 1*3 matrix XYZ
% XnYnZn is a 1*3 array representing the reference white in cie Yxy coordinates
% (C) 17/12/2001 by Yazhu Ling in University of Newcastle
% Converted for use with kccv in 21/06/2002 by Kit Wolf
function Luv = luvtoXYZ(Luv, varargin);

% prepare varargin, which has got mangled in the passing
while iscell(varargin{1}), varargin = varargin{1}; end

% search for the input argument specifying the whitepoint
if length(varargin)<2, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
for i=1:length(varargin)-1,
  if ischar(varargin{i}),
    if strcmp('whitepoint', lower(varargin{i})), XnYnZn = prime_it(kccv(varargin{i+1}, 'Yxy', 'XYZ')); ...
          break; break; end
  end
  if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
end

% Method based on formulae were originally taken from Wyszecki and Stiles,
2nd Ed. 'Color science, Concepts and Methods, Quantitative data and formulae',
% pg. 165, published by John Wiley and Sons in the USA and Canada.

% First of all, determine [Y, u', v']

% Is it possible to unambiguously determine Y/Yn from L*?
% 0.008856^0.33333333333333333333333333333*116 - 16, ans =  7.9996
% 0.008856*903.3, ans = 7.9996 - so if L < 7.9996, then this formula must have been used
tempmask = Luv(:,1) > 7.9996;

Luv(tempmask,1) = (((Luv(tempmask,1) + 16)./116).^3).*XnYnZn(1);
Luv(~tempmask,1) = (Luv(~tempmask,1)/903.3).*XnYnZn(1);

% Now determine the denominator term
denom = (Luv(:,1)*9)./Luv(:,3);

% This enables us to work out 'X'
Luv(:,2) = (Luv(:,2).*denom)/4; % Luv(:,2) now holds 'X'

% And finally, 'Z'
Luv(:,3) = (denom - Luv(:,2) - Luv(:,1)*15)./3;

% Our Luv variable now holds [Y, X, Z]; This must be rearranged to the order XYZ
denom = Luv(:,2); % use 'denom' as a swap variable
Luv(:,2) = Luv(:,1);
Luv(:,1) = denom;

% All done, and Luv now contains the variable XYZ. Sorry for any confusion
% vzeo

function primed = prime_it(XtoZ)
denom = (XtoZ(:,1) + XtoZ(:,2)*15 + XtoZ(:,3)*3);
primed = zeros(size(XtoZ));
primed(:, 1) = XtoZ(:,2);
primed(:, 2) = (XtoZ(:,1)*4)./denom; % Calculate u' (or un')
primed(:, 3) = (XtoZ(:,2)*9)./denom; % Calculate v' (or uv')
```

LUVSTARtoLHS.m

```
1  % This function convert the CIELuv values L, u and v into Luminance
2  % Hue and Saturation. Both Luv and LHS are n*3 matrix
3  % (C) Kit Wolf, University of Newcastle, 5/4/2002
4  % Formula from "Measuring colour"
by R.W.G. Hunt, 3rd edn., Fountain Press, Kingston-upon-Thames, England(1998). pg. 276
5  function LHS=LUVPRIMEtoLHS(Luv, varargin);
6
7  LHS = Luv; % allocate memory
8
9  % Do the 'L'
10 %LHS(:, 1) = Luv(:, 1);
11
12 % Do the 'H'
13 LHS(:, 2) = atan2(Luv(:,3), Luv(:,2));
14 LHS(:, 2) = LHS(:, 2) + ((LHS(:,2)<0).*2.*pi); % This line changes the range of atan2 to 0-2*pi rather than -pi to pi
15
16 % Do the 'S'
17 %LHS(:, 3) = 13*((Luv(:, 2)./(13*Luv(:,1))).^2+((Luv(:, 3)./(13*Luv(:,1)))).^2).^0.5;
18 LHS(:, 3) = ((Luv(:, 2).^2+(Luv(:, 3).^2)).^0.5)./Luv(:,1);
19
20 % Vzeo
```

LUVSTARtoXYZ.m

```
1  % This fucntion convert the CIEluv colour space into tristimulus values
2  % The luv is a n*3 matrix represent CIEL, a, b values
3  % The function return a 1*3 matrix XYZ
4  % XnYnZn is a 1*3 array representing the reference white in cie Yxy coordinates
5  % (C) 17/12/2001 by Yazhu Ling in University of Newcastle
6  % Converted for use with kccv in 21/06/2002 by Kit Wolf
7  function Luv = luvtoXYZ(Luv, varargin);
8
9  % prepare varargin, which has got mangled in the passing
10 if length(varargin) > 0, while iscell(varargin{1}), varargin = varargin{1}; end, end
11
12 % search for the input argument specifying the whitepoint
13 if ~max(strcmp(varargin, 'whitepoint')), error('A whitepoint must be specified. Format: ''whitepoint'', ...
[Y, x, y]'); end
14 for i=1:length(varargin)-1,
15   if ischar(varargin{i}),
16     if strcmp('whitepoint', lower(varargin{i})), XnYnZn = prime_it(kccv(varargin{i+1}, 'Yxy', 'XYZ')); ...
          break; break; end
17   end
18   if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
19 end
20
21 % Method based on formulae were originally taken from Wyszecki and Stiles, 2nd Ed. 'Color science,
          Concepts and Methods, Quantitative data and formulae',
22 % pg. 165, published by John Wiley and Sons in the USA and Canada.
23
24 % First of all, determine [Y, u', v']
25 Luv(:,2) = Luv(:,2)./(Luv(:,1)*13) + XnYnZn(2); % now holds u'
26 Luv(:,3) = Luv(:,3)./(Luv(:,1)*13) + XnYnZn(3); % now holds v'
27 % Is it possible to unambiguously determine Y/Yn from L*?
28 % 0.008856^0.33333333333333333333333333333*116 - 16, ans =  7.9996
29 % 0.008856*903.3, ans = 7.9996 - so if L < 7.9996, then this formula must have been used
30 tempmask = Luv(:,1) > 7.9996;
31
32 Luv(tempmask,1) = (((Luv(tempmask,1) + 16)./116).^3).*XnYnZn(1);
33 Luv(~tempmask,1) = (Luv(~tempmask,1)/903.3).*XnYnZn(1);
34
35 % Now determine the denominator term
```

```
denom = (Luv(:,1)*9)./Luv(:,3);

% This enables us to work out 'X'
Luv(:,2) = (Luv(:,2).*denom)/4; % Luv(:,2) now holds 'X'

% And finally, 'Z'
Luv(:,3) = (denom - Luv(:,2) - Luv(:,1)*15)./3;

% Our Luv variable now holds [Y, X, Z]; This must be rearranged to the order XYZ
denom = Luv(:,2); % use 'denom' as a swap variable
Luv(:,2) = Luv(:,1);
Luv(:,1) = denom;

% All done, and Luv now contains the variable XYZ. Sorry for any confusion
% vzeo

function primed = prime_it(XtoZ)
denom = (XtoZ(:,1) + XtoZ(:,2)*15 + XtoZ(:,3)*3);
primed = zeros(size(XtoZ));
primed(:, 1) = XtoZ(:,2);
primed(:, 2) = (XtoZ(:,1)*4)./denom; % Calculate u' (or un')
primed(:, 3) = (XtoZ(:,2)*9)./denom; % Calculate v' (or vn')
```

LuvtoCD.m

```
% Function returns a measure of how great the perceptual difference between two Luv specified colours is.
% Function retunrns an n*3 matrix of values, the second two columns of which are set to '0'. This is to
% ensure that all reshapes are performed correctly.
% (C) Kit WOlf, University of Newcastle upon Tyne, 2002
%
function CD = LuvtoCD(Luv, varargin);
varargin = varargin{1};
% search for the input argument specifying the whitepoint
if length(varargin)<2, error('A comparison set of colours must be specified. Format: ''ComparisonSet'', ...
        matrixname'); end
for i=1:length(varargin)-1,
  if ischar(varargin{i}),
    if strcmp('comparisonset', lower(varargin{i})), CompSet = varargin{i+1}; end
  end
%  if i==length(varargin)-1, error('A comparison set of colours must be specified. ...
Format: ''ComparisonSet'', matrixname'); end
end
CD = zeros(size(Luv));
CD(:,1) = ((CompSet(:,1)-Luv(:,1)).^2 + (CompSet(:,2)-Luv(:,2)).^2 + (CompSet(:,3)-Luv(:,3)).^2).^0.5
% Formulae taken from ''Measuring colour''
by R.W.G. Hunt, 3rd edition, Fountain press, Kingston upon Thames, UK, 1998, pg. 275
% The end
```

MCBtoXYZ.m

```
% cmat is a n*m*3 matrix of luminance, LM contrasts and S-cone contrasts; refcol = 1*1*3 vec of Yxy values
%  function im = contrast_convert(cmat, refcol)
function im = MCBtoXYZ(cmat, varargin)

% prepare varargin, which may have got mangled in the passing
while iscell(varargin{1}), varargin = varargin{1}; end

% search for the input argument specifying the whitepoint
for i=1:length(varargin)-1,
  if ischar(varargin{i}),
    if strcmp('whitepoint', lower(varargin{i})), refcol = kccv(varargin{i+1}, 'Yxy', 'XYZ'); end
  end
%  if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
end
if ~exist('refcol', 'var'), error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end

% fill the matrix with the LMS values of the whitepoint
im = zeros(size(cmat, 1), 3);
for i=1:3, im(:,i) = refcol(i); end
im = kccv(im, 'XYZ', 'LMS');

% brightness values
for i=1:3, im(:,i) = im(:,i).*(cmat(:,1)+1); end

% isoluminant L-M axis
for i=1:size(cmat, 1),
   delta = tcc(im(i, 1), im(i,2), cmat(i, 2));
   im(i, 1) = im(i, 1) + delta;
   im(i, 2) = im(i, 2) - delta;
end

% S axis
im(:,3) = im(:,3).*(1+cmat(:,3));

% reconvert to XYZ
im = kccv(im, 'LMS', 'XYZ');
```

tcc.m

```
function dL = tcc(L, M, contrast)
% function dL = tcc(L, M, contrast)
% return dL when fed baseline L & M cone excitations and desired total cone contrast - equiluminance desired
% tcc = sqrt ((dL/L)^2 + (dM/M)^2 + (dS/S)^2). In this case dS = 0, so ignored
% luminance = L + M, so dL = -dM to retain equiluminance
% equation works out as dL = dM = sqrt((contrast^2 * L^2 * M^2)/(M^2 + L^2)); (see workbook)
% if desired L-cone contrast is negative, multiply dL by sign of 'contrast' as info lost in squaring
%  ------------------------------------------------------------------------------------
%
%  KW 12/10/2001 dept. of Physiology
%
%  ------------------------------------------------------------------------------------
dL = sign(contrast)*sqrt((contrast^2*L^2*M^2)/(M^2 + L^2));
% fin
```

UI8RGBtoDBLRGB.m

```
% function dbl = UI8RGBtoDBLRGB(ui8, varargin)
% This function converts 24 bit RGB to double precision RGB, bypassing conversions to XYZ etc etc.
%
function dbl = UI8RGBtoDBLRGB(ui8, varargin)

dbl = double(ui8)/255;
```

XYZtoCC.m

```
%  cmat is a n*3 matrix of L, M, and S-cone contrasts;
%  Whitepoint must be specified.
%  refcol = 1*1*3 vector of Yxy values
%  Lnew = (1+Lcontrast)*Lold; likewise for M, S contrasts
% Kit Wolf Jan 2002, Dept. of Physiology
function cmat = XYZtoCC(im, varargin)

% prepare varargin, which may have got mangled in the passing
while iscell(varargin{1}), varargin = varargin{1}; end


% search for the input argument specifying the whitepoint
for i=1:length(varargin)-1,
  if ischar(varargin{i}),
    if strcmp('whitepoint', lower(varargin{i})), refcol = kccv(varargin{i+1}, 'Yxy', 'LMS'); end
  end
%  if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
end
if ~exist('refcol', 'var'), error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end


% do the conversion
cmat = zeros(size(im, 1), 3);

im = kccv(im, 'XYZ', 'LMS');

for i=1:3, cmat(:, i) = (im(:, i)/refcol(i))-1; end
```

XYZtoLHC.m

```
% This function convert the XYZ2LHC
% LHC represent Lightness, Hue, Chroma in CIELab colour space
% Both the output and the input are the 1*3 matrix
%
% (C) 2002, Kit Wolf, University of Newcastle upon Tyne
function cvar=XYZtoLHC(cvar, varargin) %input cvar is XYZ format
varargin = varargin{1};
cvar=XYZtoLab(cvar, varargin); %this cvar is in Lab format
cvar=LabtoLHC(cvar); % and finally, LHC, just like we want
```

XYZtoLHS.m

```
% This function converts XYZ2Luv
% Luv represent Lightness, Hue, Saturation in CIELuv colour space
% Both the output and the input are the 1*3 matrix
%
% (C) 2002, Kit Wolf, University of Newcastle upon Tyne
function cvar=XYZtoLUVPRIME(cvar, varargin) %input cvar is XYZ format
%varargin = varargin{1};
cvar=XYZtoLUVPRIME(cvar, varargin); %this cvar is in Lab format
cvar=LUVPRIMEtoLHS(cvar, varargin); % and finally, Luv, just like we want
```

XYZtoLUVPRIME.m

```
% The function convert the CIE XYZ tristimulus values into CIELuv colour space.
% The XYZ is a 1*3 matrix
% The function returns Luv (3*1 matrix with L, a, b)
% XnYnZn is a 1*3 matrix represent the tristimulus values of the reference white
% (C) 17/12/2001 by Yazhu Ling in University of Newcastle
% Update by Kit Wolf for use with kccv 21/06/2002
function XYZ = XYZtoLuv(XYZ, varargin);

% prepare varargin, which has got mangled in the passing
if length(varargin) > 0, while iscell(varargin{1}), varargin = varargin{1}; end, end

% search for the input argument specifying the whitepoint
if length(varargin)<2, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
for i=1:length(varargin)-1,
  if ischar(varargin{i}),
    if strcmp('whitepoint', lower(varargin{i})), XnYnZn = prime_it(kccv(varargin{i+1}, 'Yxy', 'XYZ')); ...
        break; break; end
  end
  if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
end

XYZ = prime_it(XYZ);

XYZ(:,1) = XYZ(:,1)./XnYnZn(:,1); % It may be called XYZ, but it now holds [Y/Yn, u', v']

% Calculate an accurate value for L*
tempmask = XYZ(:,1) > 0.008856; % a different formula must be used for very low values of XYZ
XYZ(tempmask, 1) = (XYZ(tempmask, 1).^(1/3))*116 - 16;
XYZ(~tempmask, 1) = XYZ(~tempmask, 1)*903.3;

% Formulae from Wyszecki and Stiles, 2nd Ed. 'Color science'
% pg. 165, published by John Wiley and Sons in the USA and Canada

% fin

function primed = prime_it(XtoZ)

denom = (XtoZ(:,1) + XtoZ(:,2)*15 + XtoZ(:,3)*3);
primed = zeros(size(XtoZ));
primed(:, 1) = XtoZ(:,2);
primed(:, 2) = (XtoZ(:,1)*4)./denom; % Calculate u' (or un')
primed(:, 3) = (XtoZ(:,2)*9)./denom; % Calculate v' (or vn')
```

XYZtoLUVSTAR.m

```
% The function convert the CIE XYZ tristimulus values into CIELuv colour space.
% The XYZ is a 1*3 matrix
% The function returns Luv (3*1 matrix with L, a, b)
% XnYnZn is a 1*3 matrix represent the tristimulus values of the reference white
% (C) 17/12/2001 by Yazhu Ling in University of Newcastle
% Update by Kit Wolf for use with kccv 21/06/2002
function XYZ = XYZtoLuv(XYZ, varargin);

% prepare varargin, which has got mangled in the passing
if length(varargin) > 0, while iscell(varargin{1}), varargin = varargin{1}; end, end

% search for the input argument specifying the whitepoint
if ~strcmp(varargin, 'whitepoint'), error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
for i=1:length(varargin)-1,
  if ischar(varargin{i}),
```

```
16     if strcmp('whitepoint', lower(varargin{i})), XnYnZn = prime_it(kccv(varargin{i+1}, 'Yxy', 'XYZ')); ...
           break; break; end
17    end
18    if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
19  end
20
21  XYZ = prime_it(XYZ);
22
23  XYZ(:,1) = XYZ(:,1)./XnYnZn(:,1); % It may be called XYZ, but it now holds [Y/Yn, u', v']
24
25  % Calculate an accurate value for L*
26  tempmask = XYZ(:,1) > 0.008856; % a different formula must be used for very low values of XYZ
27  XYZ(tempmask, 1) = (XYZ(tempmask, 1).^(1/3))*116 - 16;
28  XYZ(~tempmask, 1) = XYZ(~tempmask, 1)*903.3;
29
30  % Calculate u*
31  XYZ(:,2) = 13*XYZ(:,1).*(XYZ(:,2)-XnYnZn(:, 2)); % XYZ(:,2) now holds u*
32
33  % Calculate v*
34  XYZ(:,3) = 13*XYZ(:,1).*(XYZ(:,3)-XnYnZn(:, 3)); % XYZ(:,2) now holds v*
35
36
37  % XYZ now contains L*u*v* data
38  % Formulae from Wyszecki and Stiles, 2nd Ed. 'Color science'
39  % pg. 165, published by John Wiley and Sons in the USA and Canada
40
41  % fin
42
43  function primed = prime_it(XtoZ)
44
45  denom = (XtoZ(:,1) + XtoZ(:,2)*15 + XtoZ(:,3)*3);
46  primed = zeros(size(XtoZ));
47  primed(:, 1) = XtoZ(:,2);
48  primed(:, 2) = (XtoZ(:,1)*4)./denom; % Calculate u' (or un')
49  primed(:, 3) = (XtoZ(:,2)*9)./denom; % Calculate v' (or vn')
```

XYZtoLab.m

```
1   % The function converts CIE XYZ tristimulus values into CIELab colour space.
2   % The XYZ is a 1*3 matrix
3   % The function returns Lab (3*1 matrix with L, a, b)
4   % XnYnZn is a 1*3 matrix represent the tristimulus values of the reference white
5   % (C) 17/12/2001 by Yazhu Ling in University of Newcastle
6   % Update by Kit Wolf for use with kccv 21/06/2002
7   function Lab = XYZtoLab(XYZ, varargin);
8
9   % prepare varargin, which has got mangled in the passing
10  varargin = varargin(1);
11  varargin = varargin{1};
12
13  % search for the input argument specifying the whitepoint
14  if length(varargin)<2, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
15  for i=1:length(varargin)-1,
16    if ischar(varargin{i}),
17      if strcmp('whitepoint', lower(varargin{i})), XnYnZn = kccv(varargin{i+1}, 'Yxy', 'XYZ'); ...
break; break; end
18    end
19    if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
20  end
21  %
22  sz = size(XYZ);
23  conv_I = zeros(sz); % Allocate memory
24  conv_F = zeros(sz); % Allocate memory
25  Lab = zeros(sz); % Allocate memory
26  clear sz
```

```
27
28  for i=1:3,
29    conv_I(:, i) = XYZ(:, i) ./ XnYnZn(i);
30    conv_F(:, i) = (conv_I(:, i)>=0.008856).*conv_I(:, i).^(1/3) + (conv_I(:, i)<0.008856).*...
(conv_I(:, i)*(7.7871+(16/116)));
31  end
32
33  % Formulae taken from "Measuring colour"
by R.W.G. Hunt, 3rd edition, Fountain press, Kingston upon Thames, UK, 1998, pg. 275
34  Lab(:, 1) = 116 * conv_F(:, 2)-16;
35  Lab(:, 2) = 500 * (conv_F(:, 1)-conv_F(:, 2));
36  Lab(:, 3) = 200 * (conv_F(:, 2)-conv_F(:, 3));
37
38  % fin
```

XYZtoMCB.m

```
1   % cmat is a n*m*3 matrix of luminance, LM contrasts and S-cone contrasts; refcol = 1*1*3 vec of Yxy values
2   %  function im = contrast_convert(cmat, refcol)
3   function cmat = XYZtoMCB(im, varargin)
4
5   % prepare varargin, which may have got mangled in the passing
6   while iscell(varargin{1}), varargin = varargin{1}; end
7
8   % search for the input argument specifying the whitepoint
9   for i=1:length(varargin)-1,
10    if ischar(varargin{i}),
11      if strcmp('whitepoint', lower(varargin{i})), refcol = kccv(varargin{i+1}, 'Yxy', 'LMS'); end
12    end
13  %  if i==length(varargin)-1, error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
14  end
15  if ~exist('refcol', 'var'), error('A whitepoint must be specified. Format: ''whitepoint'', [Y, x, y]'); end
16
17  % allocate storage space for the output variable
18
19  cmat = zeros(size(im));
20
21  %for i=1:3, cmat(:,i) = refcol(i); end
22
23  % reconvert to LMS
24  im = kccv(im, 'XYZ', 'LMS');
25
26  % rescale for brightness
27  cmat(:,1) = (im(:,1)+im(:,2))./(refcol(1)+refcol(2));
28  for i=1:3, im(:, i) = im(:, i)./cmat(:,1); end
29  cmat(:,1) = cmat(:,1) - 1;
30
31  % LM axis
32  delta = im(:,1) - refcol(1);
33  cmat(:,2) = sign(delta).*(sqrt((delta/refcol(1)).^2 + (delta/refcol(2)).^2));
34
35  % S axis
36  cmat(:,3) = im(:,3)/refcol(3) - 1;
```

XYZtoYxy.m

```
1  function Yxyim = XYZ2Yxyim(XYZim, varargin)
2  % converts matrix of Yxy values into matrix of XYZ values
3  % takes as argument matrix m*n*3 matrix, (a, b, 1) = X; 2nd = (a, b, 2) = Y; 3rd = (a, b, 3) = Z...
4  %
5  % allocate memory for return array
6  Yxyim = zeros(size(XYZim));
7  %
8  % Y value unchanged by conversion - only position to be changed
9  Yxyim(:, 1) = XYZim(:, 2);
10 %
11 % work out (X + Y + Z) for each element
12 XYZsum = sum(XYZim, 2);
13 % x = X / (X + Y + Z);
14 Yxyim(:, 2) = XYZim(:, 1)./XYZsum;
15 %
16 % y = Y / (X + Y + Z);
17 Yxyim(:, 3) = XYZim(:, 2)./XYZsum;
18 % end of function
```

YxytoXYZ.m

```
1  function XYZim = Yxy2XYZim(Yxyim, varargin)
2  % converts matrix of Yxy values into matrix of XYZ values
3  % takes as argument matrix m*3 matrix, (m, 1) = Y; 2nd = (a, 2) = x; 3rd = (a, 3) = y...
4  %
5  % allocate memory for return array
6  XYZim = zeros(size(Yxyim));
7  %
8  % Y (luminance) term is passed unchanged
9  XYZim(:, 2) = Yxyim(:, 1);
10 %
11 % X = (x.(X+Y+Z)/(y.(X+Y+Z)).Y = (x/y).Y
12 XYZim(:, 1) = (Yxyim(:, 2)./Yxyim(:, 3)).*Yxyim(:, 1);
13 %
14 % Z = Y.(1-y-x)/y
15 XYZim(:, 3) = Yxyim(:, 1).*((1 - Yxyim(:, 2) - Yxyim(:, 3)))./Yxyim(:, 3);
16 % end of function
```

high_freq_test.m

```
1  % function high_freq_test(orientation, pixels, gun)
2  % orientation - 1 == horizontal; anything else == vertical
3  % pixels - how wide should each line be?
4  % gun - 123 == rgb
5  function high_freq_test(orientation, pixels, gun)
6
7  pixels = pixels*2;
8
9  % position figure
10 startpos = get(0, 'ScreenSize')/2-100; startpos = startpos(3:4);
11 set(gcf, 'position', [startpos, 200, 200], 'menubar', 'none');
12 set(gca, 'position', [0, 0, 1, 1]);
13
14 im = (mod(meshgrid(1:200, 1:200), pixels) + 1 > (pixels/2)).*gun+1;
15 cmap = [0, 0, 0; 1, 0, 0; 0, 1, 0; 0, 0, 1]
16
17 % draw the stimulus
18 if orientation == 1, image(im'); else image(im); end
19 colormap(cmap);
20
21 figure(gcf)
```

my_ginput.m

```
1  function [out1,out2,out3] = ginput(arg1)
2  %GINPUT Graphical input from mouse.
3  %  [X,Y] = GINPUT(N) gets N points from the current axes and returns
4  %  the X- and Y-coordinates in length N vectors X and Y. The cursor
5  %  can be positioned using a mouse (or by using the Arrow Keys on some
6  %  systems). Data points are entered by pressing a mouse button
7  %  or any key on the keyboard except carriage return, which terminates
8  %  the input before N points are entered.
9  %
10 %  [X,Y] = GINPUT gathers an unlimited number of points until the
11 %  return key is pressed.
12 %
13 %  [X,Y,BUTTON] = GINPUT(N) returns a third result, BUTTON, that
14 %  contains a vector of integers specifying which mouse button was
15 %  used (1,2,3 from left) or ASCII numbers if a key on the keyboard
16 %  was used.
17
18 %  Copyright 1984-2001 The MathWorks, Inc.
19 %  $Revision: 5.30 $ $Date: 2001/04/15 12:03:29 $
20
21 out1 = []; out2 = []; out3 = []; y = [];
22 c = computer;
23 if ~strcmp(c(1:2),'PC') & ~strcmp(c(1:2),'MA')
24   tp = get(0,'TerminalProtocol');
25 else
26   tp = 'micro';
27 end
28
29 if ~strcmp(tp,'none') & ~strcmp(tp,'x') & ~strcmp(tp,'micro'),
30   if nargout == 1,
31    if nargin == 1,
32      out1 = trmginput(arg1);
33    else
34      out1 = trmginput;
35    end
36   elseif nargout == 2 | nargout == 0,
```

```
37    if nargin == 1,
38      [out1,out2] = trmginput(arg1);
39    else
40      [out1,out2] = trmginput;
41    end
42    if nargout == 0
43      out1 = [ out1 out2 ];
44    end
45   elseif nargout == 3,
46    if nargin == 1,
47      [out1,out2,out3] = trmginput(arg1);
48    else
49      [out1,out2,out3] = trmginput;
50    end
51   end
52  else
53
54   fig = gcf;
55   figure(gcf);
56
57   if nargin == 0
58    how_many = -1;
59    b = [];
60   else
61    how_many = arg1;
62    b = [];
63    if isstr(how_many) ...
64       | size(how_many,1) ~= 1 | size(how_many,2) ~= 1 ...
65       | ~(fix(how_many) == how_many) ...
66       | how_many < 0
67      error('Requires a positive integer.')
68    end
69    if how_many == 0
70      ptr_fig = 0;
71      while(ptr_fig ~= fig)
72       ptr_fig = get(0,'PointerWindow');
73      end
74      scrn_pt = get(0,'PointerLocation');
75      loc = get(fig,'Position');
76      pt = [scrn_pt(1) - loc(1), scrn_pt(2) - loc(2)];
77      out1 = pt(1); y = pt(2);
78    elseif how_many < 0
79      error('Argument must be a positive integer.')
80    end
81   end
82
83   % Remove figure button functions
84   state = uisuspend(fig);
85   pointer = get(gcf,'pointer');
86
87  % My modifications (Kit). Make the pointer transparent (disappear).
88  % I can't find a better way of doing it, alas.
89  a = zeros(16);
90  a(:,:) = nan;
91  set(gcf, 'pointershapeCdata', a, 'pointer', 'custom');
92  %  set(gcf,'pointer','fullcrosshair');
93   fig_units = get(fig,'units');
94   char = 0;
95
96   while how_many ~= 0
97    % Use no-side effect WAITFORBUTTONPRESS
98    waserr = 0;
99    try
100         keydown = wfbp;
101    catch
```

```
102         waserr = 1;
103       end
104       if(waserr == 1)
105        if(ishandle(fig))
106          set(fig,'units',fig_units);
107             uirestore(state);
108          error('Interrupted');
109        else
110          error('Interrupted by figure deletion');
111        end
112       end
113
114       ptr_fig = get(0,'CurrentFigure');
115       if(ptr_fig == fig)
116        if keydown
117          char = get(fig, 'CurrentCharacter');
118          button = abs(get(fig, 'CurrentCharacter'));
119          scrn_pt = get(0, 'PointerLocation');
120          set(fig,'units','pixels')
121          loc = get(fig, 'Position');
122          pt = [scrn_pt(1) - loc(1), scrn_pt(2) - loc(2)];
123          set(fig,'CurrentPoint',pt);
124        else
125          button = get(fig, 'SelectionType');
126          if strcmp(button,'open')
127           button = b(length(b));
128          elseif strcmp(button,'normal')
129           button = 1;
130          elseif strcmp(button,'extend')
131           button = 2;
132          elseif strcmp(button,'alt')
133           button = 3;
134          else
135           error('Invalid mouse selection.')
136          end
137        end
138        pt = get(gca, 'CurrentPoint');
139
140        how_many = how_many - 1;
141
142        if(char == 13) % & how_many ~= 0)
143          % if the return key was pressed, char will == 13,
144          % and that's our signal to break out of here whether
145          % or not we have collected all the requested data
146          % points.
147          % If this was an early breakout, don't include
148          % the <Return> key info in the return arrays.
149          % We will no longer count it if it's the last input.
150          break;
151        end
152
153        out1 = [out1;pt(1,1)];
154        y = [y;pt(1,2)];
155        b = [b;button];
156       end
157      end
158
159      uirestore(state);
160      set(fig,'units',fig_units);
161
162      if nargout > 1
163       out2 = y;
164       if nargout > 2
165        out3 = b;
166       end
```

```
167   else
168     out1 = [out1 y];
169   end
170
171  end
172
173  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
174  function key = wfbp
175  %WFBP  Replacement for WAITFORBUTTONPRESS that has no side effects.
176
177  fig = gcf;
178  current_char = [];
179
180  % Now wait for that buttonpress, and check for error conditions
181  waserr = 0;
182  try
183   h=findall(fig,'type','uimenu','accel','C');  % Disabling ^C for edit menu so the only ^C is for
184   set(h,'accel','');             % interrupting the function.
185   keydown = waitforbuttonpress;
186   current_char = double(get(fig,'CurrentCharacter')); % Capturing the character.
187   if~isempty(current_char) & (keydown == 1)      % If the character was generated by the
188           if(current_char == 3)          % current keypress AND is ^C, set 'waserr'to 1
189                       waserr = 1;             % so that it errors out.
190           end
191   end
192
193   set(h,'accel','C');               % Set back the accelerator for edit menu.
194  catch
195   waserr = 1;
196  end
197  drawnow;
198  if(waserr == 1)
199   set(h,'accel','C');              % Set back the accelerator if it errored out.
200   error('Interrupted');
201  end
202
203  if nargout>0, key = keydown; end
```

pixelsize.m

```
1   function pixels = pixelsize(degrees)
2   %  Using information in monitordef.mat on size of the screen, resolution and viewing distance to calculate the
number of pixels
3   %  corresponding to argument 'degrees'. i.e. a square pixels by pixels will cover 'degrees' of visual field when
viewed under the conditions
4   %  specified in monitordef.mat
5   %  ----------------------------------------------------------------------------
6   %
7   %  KW 11/10/2001 University of Newcastle upon Tyne, dept. of physiology
8   %
9   %
10  %
11  %  ----------------------------------------------------------------------------
12  global MonitorDat
13
14  pixels = round(degrees * MonitorDat.pd);
15                                          %                   end                   of                   function
```

Appendix D – A Brief Survey of the Use of Inferential Statistics in Psychophysics Experiments

Introduction

When asked to include additional statistical analyses in my thesis, I was initially very surprised. Hypothesis-testing is almost ubiquitous within psychology as a whole – a literature survey of leading psychology journals showed that over 95% of experimental papers published between 1998 and 2006 cite p-values (Cumming and McMenamin). Yet even when surveying the literature prior to undertaking my research, it had been a cause of some curiosity that the statistics revolution did not appear to have reached the field of psychophysics. Robert Boynton expressed what he felt to be the status-quo in 1993: "*Experience has shown that [when highly trained and motivated observers are used], important relations show up with such strength that statistical tests are not even needed*" (White and White 1995).

Looking over the papers that I read before starting my experiments, I noticed that those from the mid-1970s and later generally provided only error bars, whereas earlier papers provided few if any statistics – e.g. Jameson & Hurvich estimate a single confidence-interval in their 1960 paper 'Opponent Chromatic Induction' (Jameson and Hurvich 1960) and neither Stewart's 1959 paper on the Gelb effect (Stewart 1959) nor Walraven's 1973 paper on stray light artifacts (Walraven 1973), nor Valberg's 1974 paper on induction (Valberg 1974) employ any means of hypothesis testing or confidence-intervals whatsoever. Moving to more recent papers, a number that are central to my research quote no statistics: Webster and Mollon's 1994 paper plots datapoints that are the means of either 3 or 6 adjustment matches and calculates least-squares fits, but provides no confidence intervals for them – or indeed any further statistics (Webster and Mollon 1994). Neither do Zaidi et al's 1991 (Zaidi, Yoshimi et al. 1991) or 1992 (Zaidi, Yoshimi et al. 1992) papers, which also used a nulling adjustment paradigm – this time with 30 or 10 trials per point respectively. Jenness and Shevell (Jenness and Shevell 1995) and Barnes and Shevell (Barnes and Shevell 2002) also use an adjustment paradigm and show error bars, but again do not calculate any p-values. Brown and MacLeod also employ an adjustment technique, and show error bars but do not report p-values (Brown and MacLeod 1997). Despite specifically comparing different models of visual processing, Chichilnisky and Wandell reported neither error-bars, nor p-values, nor other statistical tests (Chichilnisky and Wandell 1995) such as calculating Akaike's information-criterion (Akaike 1974). The notable exceptions are Wachtler, Albright and Sejnowski (Wachtler, Albright et al. 2001), who in 2001 report confidence intervals and p-values calculated using the chi-squared statistic – but only for selected pairwise comparisons, and Shevell and Miller (Shevell and Miller 1996) and Shevell and Wei (Shevell and Wei 2000) and. In both of the latter experiments, methods of adjustment were employed whereas Wachtler et al. employed the method of constant stimuli.

This illustrates why I was surprised to be asked to include hypothesis tests – they were rarely used in the psychological literature relevant to this thesis. But it does not exclude the possibility that they are more commonly used in other areas. Could significance testing be more widely used in other fields of psychophysics research? If so, was this also true when I undertook my PhD research? And what statistical techniques might be used to analyse method-of-constants experiments?

To determine the extent to which statistics methods are used in other areas of visual psychophysics, I decided to perform a further search through a number of recent psychophysics papers. I took the last volume of Vision Research, 1995, as a starting point. I chose this journal because it contains a high proportion of psychophysics papers, and unlike some other journals (e.g. Nature Neuroscience) it is divided into a number of different sections which makes papers based on psychophysics easier to find. It is held in physical form at the local University library and is therefore more readily searchable than internet holdings. I also conducted a search of the last 2003 volume of Vision Research to determine whether these statistical tests are now used more commonly. The year 2003 was chosen simply because it is the last date on which the papers were still divided into separate categories in the bound volumes.

Whilst searching through the issues, I originally aimed only to examine papers that were in the 'Psychophysics' sections. I also examined dispatches if they appeared to use psychophysics methods, and in some cases I may have inadvertently included papers in the 'Behavioural physiology and visuomotor control' section as the dividing pages sometimes appeared to be missing from the bound volumes, and as these papers frequently employed psychophysics methods. I excluded reviews, and only classified papers by whether or not they cited p-values and according to the method used to measure the psychometric function (Pelli and Farell 1999), whether it be a 'method of constants', 'adaptive psychometric method' such as QUEST (Watson and Pelli 1983) or ZEST (Kingsmith, Grigsby et al. 1994), or an adjustment technique, or an 'other' technique that was not straightforward to classify. I do not think I included any inappropriate papers, so this should not affect my conclusions. Only one paper used more than one method, but again this did not affect its classification.

Use of statistics in 1995

I examined a total of 22 papers from 1995, of which five were unclassifiable. It was immediately obvious that many papers in 1995 did not include any statistical analyses other than error bars. Indeed, one paper did not even include error-bars. Overall, fifteen papers did not quote p-values, and seven did. Subdividing the papers based on the psychophysical method used, it's notable that five papers used the method of constant stimuli, and that none of these quoted any p-values. Conversely, seven of twelve papers based on adaptive psychometric methods or methods of adjustment methods quoted p-values and five did not.

Use of statistics in 2003

In 2003, only 9 out of 27 papers surveyed did not quote any p-values, and many of these used unusual methods, or quoted other statistics such as regression coefficients or values for d'. Again, a single paper neither quoted p-values nor displayed error bars. Papers using either an adaptive method or matching-by-adjustment were overwhelmingly likely to quote p-values – only one of twelve papers using these methods did not. However, the most pertinent difference was that both the papers based on the 'method of constant stimuli' included p-values, whereas in 1995 none of the five papers using this method had done so. Yet although both these papers stated that ANOVA had been employed, neither gave sufficient information for me to be absolutely certain how the technique had been applied to the data.

Discussion

Over the past decade, there has been a marked change in the use of statistics in psychophysics papers published in Vision Research. 2003 papers are significantly more likely to quote p-values than 1995 papers ($\boldsymbol{p} = 0.007$), and this holds true even for papers using the method of constants ($p = 0.048$), or adaptive or adjustment methods ($p=0.155$) – although the latter comparison failed to attain statistical significance. All p-values in this paragraph were determined using Fisher's exact probability test – two tailed.

Why should this be? Could it be that appropriate statistical methods did not exist in 1995? This appears unlikely. It is true that statistical techniques for the analysis of psychometric data are still undergoing rapid development, in large part because they appear well-suited to analysis by resampling methods such as Monte-Carlo or bootstrap methods, which are very computationally intensive. The curve-fitting algorithm that I used in my first experiments used Chi-squared analysis to produce its error bars and gave results within a matter of seconds; Wichmann's 'pfit' package can take several minutes to complete, even though it runs on a much faster computer. It seems reasonable to conclude that the newer Monte-Carlo methods would have been very time-consuming and unwieldy even just a few years ago, when computing power was more expensive than at present.

However, when the ‘method of constants’ experiments are subjected to further scrutiny, the technique used for significance testing generally appears to be ANOVA or a closely related technique such as repeated-measures ANOVA. These techniques are much older – ANOVA was described in Fisher’s 1925 book ‘Statistical methods for research workers’ (Fisher 1928). If it was not used in 1995 then it seems likely that this was because it was not considered necessary or simply because there was no tradition of doing so. Although ANOVA-like methods do not necessarily employ F-statistics, neither of the 2003 papers discussed any modifications or corrections to the standard ANOVA procedure.

An alternative explanation is that the journal has changed its policy to encourage the use of statistical tests. However, the current guidelines for submission do not reflect this (Editor 2008), though it may be an informal policy on the part of editors and reviewers. Even in 1995 the majority of papers using either adaptive psychometric methods or methods of adjustment cited p-values, whereas none of the papers using the ‘method of constants’ did so. Both methods of adjustment and adaptive psychometric methods provide repeated, continuous measures of psychometric thresholds which are intuitively better suited to further analysis by ANOVA than the binomial results produced by the ‘method of constants’. It therefore seems likely that in 1995 most psychophysicists were already willing to use formal tests of statistical significance for types of data for which they seemed appropriate.

Search engine statistics

Another method of addressing the problem is to use a search engine such as Google-scholar to examine trends in the use of statistics in psychophysics papers over time, as illustrated in figure 69. The graph was generated by searching for the terms ‘psychophysics’ AND ‘$p<0.05$’ in papers published over 5-year blocks of time ranging from 1970 to the present, and dividing the number of papers found by the number of papers including the term ‘psychophysics’. Although this method is fraught with sources of error, the trend is unequivocal: inferential statistics are used far more frequently today than was the case even a decade ago. This remains true even when the search terms are changed slightly – for example if ‘p-value’ or ‘$p<0.01$’ is used in the place of the term ‘$p<0.05$’ (not shown).

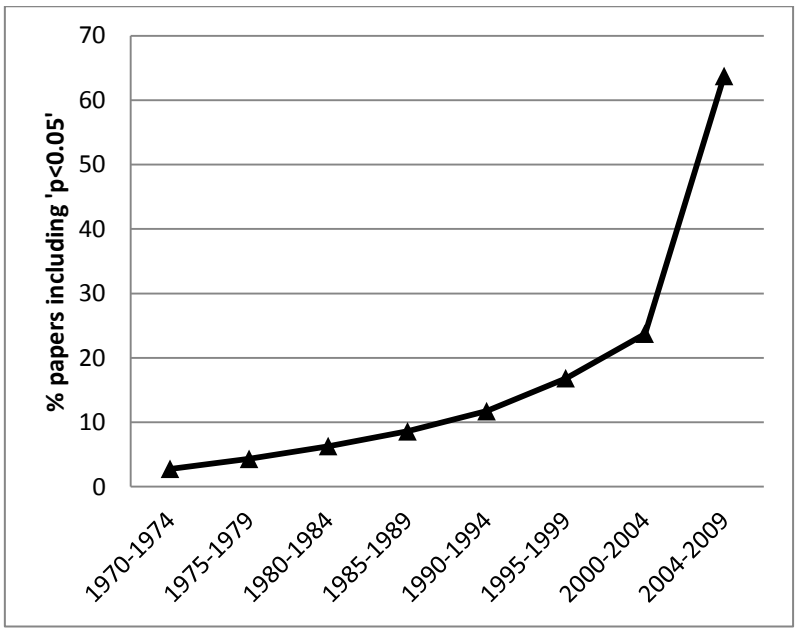


Figure 69: trends in the use of p-values in psychophysics papers, determined using 'Google scholar'. Each datapoint was determined by dividing the number of papers utilising the search terms 'psychophysics' and 'p<0.05' by the number of papers that included the search term 'psychophysics', further subdivided into the time-blocks shown on the x-axis. Although the absolute proportion of papers is certainly inaccurate as some papers may use terms such as 'p=0.03' and therefore not be counted, and papers employing psychophysics methods may never use the term 'psychophysics', the general trend remains clear: inferential statistical tests are now used far more commonly than was the case even a decade ago. The trend remains robust, even if the terms 'p-value' or 'p<0.01' are substituted for 'p<0.05'.

Year	'Psychophysics' + 'p<0.05'	'Psychophysics' only	Percentage
1970-1974	157	5650	2.7%
1975-1979	302	6980	4.3%
1980-1984	564	8960	6.3%
1985-1989	903	10500	8.6%
1990-1994	1620	13800	11.7%
1995-1999	3400	20200	16.8%
2000-2004	5580	23500	23.7%
2005-present	10200	16000	63.8%

Table 37: the data used to draw figure 69. The numbers in the middle two columns represent the number of 'hits' obtained by Google scholar when used to search for the search terms quoted at the top cell, during the dates specified in the left hand column. The searches were performed in August 2009. The trend shows that the use of statistics in psychophysics is increasing, but should not be used to indicate that '64% of psychophysics papers now employ inferential statistics' as the search may include papers from other fields that refer to psychophysics.

Conclusions

This survey is clearly narrow in its scope. Papers published in 'Vision Research' cannot be taken to be representative of papers published in the field as a whole, as editorial or other factors could have strongly influenced the findings described here. It would be interesting – though time-consuming – to extend this investigation to include psychophysics papers published in other journals, and perhaps in other areas of vision-research. However, the survey is sufficient to show that significance testing was *not* ubiquitous in psychophysics research around the time that I started my experiments, even in a reasonably high-impact journal such as Vision Research (impact factor calculated in 2008 = 2.055 (Elsevier 2008)).

Although the method is relatively crude, the results of searches carried out on 'Google scholar' broadly confirm this general impression, also widening the scope of investigation to include psychophysics papers that are not based in the field of vision-research.

Regarding the second question: whether I could find any detailed guidelines on how to apply ANOVA or other techniques to generate p-values for method-of-constants data, the answer is 'no'. However, it is impossible to prove that none exist; only to expand one's search indefinitely, or to decide at a certain point that the requirements of 'due-diligence' have been met and move on. Following a fruitless web-of-science search and consultation with a statistician, I decided to compare four potential approaches to the significance testing of psychometric functions, testing them on simulated datasets. These investigations are described in the next section, Appendix E.

Appendix E - Statistical Tests on Continuous and Categorical Data

The underlying concepts of null hypothesis significance testing (NHST, as discussed in chapter 3 - Methods) are supported by a dizzying range of statistical tests designed to work in different scenarios, but the most commonly used statistical techniques such as ANOVA are designed for use on continuous, normally distributed data. In this thesis, only the third set of experiments that use a nulling technique contain any results that are in this form, and the specific use of ANOVA and normality testing are discussed in detail in that chapter.

Unfortunately, psychometric tests often produce raw data in the form of categorical data – whether these data are in the form of binomial responses as with a 2-alternative-forced-choice test, or in the form of multivariate or categorical responses as with a four-alternative-forced-choice technique. The situation is further complicated by the fact that these responses are often not analysed directly, but are first fit by psychometric functions. Although many techniques exist for the statistical analysis of multivariate data, for this reason it is often not always obvious how to apply them to psychophysics data.

A survey of textbooks of all the graduate-level textbooks of psychophysics and psychology in the university library found no discussion of any pertinent techniques (Marks 1974; Green and Swets 1988; Gescheider 1997; Levine and Levine 2000; Breakwell 2006). Although most mentioned the possibility of using maximum-likelihood methods to fit psychometric functions to the raw data, none mentioned any form of statistical analysis more advanced than using Chi-Squared analysis to estimate confidence-intervals for the calculated parameters (Green and Swets 1988). A selection of general textbooks on ANOVA and other hypothesis-testing techniques were similarly unhelpful (Robson 1994; Hopkins, Hopkins et al. 1996; Cohen 2007). I surveyed a much larger selection of textbooks on statistics, aimed at both undergraduates and postgraduates, and found no discussions of how to use inferential statistics to compare psychometric functions.

The only explicit discussions of the problem that I found were in an introduction to the somewhat limited Monte-Carlo procedure 'pfcmp', where Hill mentions only two methods of applying NHST to psychophysics – 'pfcmp', and that of making 'an informal judgment based on whether the error-bars overlap' (Hill 2002-2008). Wachtler and Sejnowski (Wachtler, Albright et al. 2001) report p-values calculated using the chi-squared statistic – but only for selected pairwise comparisons. Might the technique be expanded? In a discussion of fitting models to continuous data – a similar problem where the F-statistic is used in place of the Chi-Squared statistic, Motulsky and Christopoulos state: "The F-test cannot be readily extended to compare three or more models. The problem is one of interpreting multiple p-values (Motulsky and Christopolous 2003)". My understanding is that this limitation would be equally applicable to calculations made on the basis of the Chi-Squared statistic.

Significance-testing and the method of constant stimuli

Here I discuss four approaches to significance-testing of data produced by the method of constant stimuli.

1. Comparisons of error bars
2. Monte-Carlo analyses
3. ANOVA based on inter-session variances
4. Logistic regression

Comparing error bars

Error bars are widely used in graphs, to give an intuitive impression of the degree of uncertainty inherent in the reporting of any value. They are widely believed to indicate whether datapoints are statistically significantly different from one another – and indeed are sometimes used for this purpose. However in most cases these comparisons are not valid, for two main reasons:

1. If two error bars showing 95% confidence intervals overlap, it does not necessarily mean that the two datapoints are not statistically significantly different from one another at the 95% level.

2. If a large number of comparisons are being made, even if the error bars do not overlap this does not necessarily mean that any two given datapoints are statistically significantly different from one another.

Methods do exist, whereby error bars may be adjusted so that comparisons of overlap can give reliable data about whether or not two or more datapoints are statistically significantly different. For example, to judge whether two datapoints are statistically significant at the 95% level, the error bars for both datapoints should be calculated by multiplying the standard error of the mean by 1.39 (Goldstein and Healy 1995). In contrast, the sizes of 95% confidence-intervals are calculated by multiplying the standard error of the mean by 1.96. Similar corrections may be calculated so that any specified number of comparisons may be made (Goldstein and Healy 1995; Ludbrook 2000).

One drawback of this technique is that it does not report absolute p-values, merely whether or not the desired level of significance has been attained. Even whilst remembering that effect-size is only indirectly related to the p-value, it is possible to say that a very low or relatively high p-value may allow us to be more or less confident in a result, even if both reach statistical significance.

Ultimately I chose not to use this technique, in part because it is unfamiliar to most people. Error bars customarily indicate either 95% confidence intervals or the standard error of the mean, and to depart from these conventions could be confusing.

Further limitations to the technique of using error bars to determine statistical significance are that it does not enable more complex tests to be made, in the same way that a paired T-test or 2-way ANOVA enables data from different observers to be combined for analysis. It can also be difficult to predetermine the number of comparisons that a reader might wish to make, and I believe that this is the case for much of my data. If a reader made any additional comparisons beyond those originally intended, he or she would be in danger of making a type-I statistical error.

Despite the fact that I do not use error bars to judge significance, I do report them in nearly all cases in order to illustrate confidence intervals for my data. Error bars may – and probably should – be shown whenever possible (APA 2001), even if alternative methods have been used to calculate statistical significance.

Monte-Carlo analyses – pfcmp

The program 'pfcmp' forms part of the 'psignifit' toolbox provided by Wichmann and Hill (Hill 1996-2009; Hill 2002-2008). It allows the comparison of two psychometric functions measured using the method of constants, using a Monte-Carlo simulation. As such, it tests the hypothesis that they were drawn from the same distribution, not that they have the same thresholds or slopes. In other words, two curves with differing slopes but an identical threshold might be judged different. In practice, it is important to know in what respects two psychometric functions differ, as depending on the experiment different aspects of the psychometric function may be most interesting. Throughout this thesis, I am more interested in thresholds than in the slope of the data so for my purposes, the possibility that a significant result may be due to a difference in slope rather than a difference in threshold, is a drawback.

As 'pfcmp' is a Monte-Carlo technique, it can produce different p-values each time it is run. However, their variance is determined by the number of repeats that are chosen, and can therefore be made arbitrarily small. For example, for observed p=0.05, the accuracy from 10,000 random permutations is approximately +/- 0.0056, determined using a binomial 99% confidence interval. I ran 10,000 simulations on each occasion that I cite p-values determined using 'pfcmp'.

A second disadvantage to the 'pfcmp' function is that it is limited to comparisons of only two datasets at a time. This may be countered by carefully choosing a small number of separate comparisons, such that if they are found to be true it can be said that the experiment shows a significant result. However, if multiple independent comparisons are made, then the probability of making a type-I statistical error also increases. A multiple-comparison method must therefore be applied to adjust the threshold alpha value accordingly.

I used a Bonferroni correction on each occasion that I used the 'pfcmp' technique. This is applied simply by counting the number of comparisons to be made and dividing the threshold value of α by this number. For example, if 5 comparisons are to be made using the customary $\alpha=0.05$ level of significance, then the target level of α should be $0.05/5 = 0.01$. The Bonferroni technique is considered to be quite conservative and has very little power if the number of comparisons is large. This causes the possibility of a 'type I error' to decrease, at the expense of increasing the risk of making a 'type II' error.

Performing ANOVA using inter-session or inter-block variance

ANOVA, or 'analysis of variance', is a method designed for the analysis of continuous, rather than categorical data, and as such cannot be applied directly to raw data obtained from 'method of constants' experiments. However, after a psychometric function has been fit to the data, the estimates for the threshold appear to be continuous, and one might therefore expect that they could be investigated using standard ANOVA techniques.

On the other hand, the thresholds calculated by a fitting program are the outputs of a complex mathematical function, and differ in important respects from the continuous data to which ANOVA is normally applied. Most obviously, threshold estimates are accompanied by a measure of their variance in the form of a confidence-interval whereas classical measurements are not (e.g. the height of an individual person picked at random off the street will be a single figure). If I take several thresholds, discarding the confidence-interval on each occasion, then I will eventually be able to indirectly recapture an estimate of the underlying variance. However, this approach appears to waste a significant amount of information. The question here is one of efficiency in the use of information, or in statistical terminology, 'power'. Might we predict that my p-values would be erroneously high, considering the amount of data collected?

Another set of concerns relates to the distribution of values that may be obtained by the function. Although the thresholds are expressed on a continuous scale, they are not truly continuous and can only take on a set number of values. Although pfit uses a bootstrapping technique to estimate variance, the maximum-likelihood technique used to estimate the thresholds is deterministic and if the same data are re-analysed the same answer is obtained. At a first glance, this may not appear to be an issue: for a reasonable-length block of five trials per-point, sampled at six points, then there are 6^5 possible combinations or 7776 possible threshold estimates. Yet the practical number of combinations is far smaller. It is customary to choose at least one stimulus level at which the observer almost always performs at 0%, and another where the observer responds at 100%. This aids in determining an accurate lapse rate for the psychometric function, but leaves only leaves four remaining levels – or 1024 possible values. For individual levels, some values are more likely than others (if you toss a coin five times, you are more likely to throw tails twice than five times) and some datapoints contribute more to the estimate of the threshold than others. The end result is that the data become quantised, and are truly semi-continuous rather than continuous (Velleman and Wilkinson 1993).

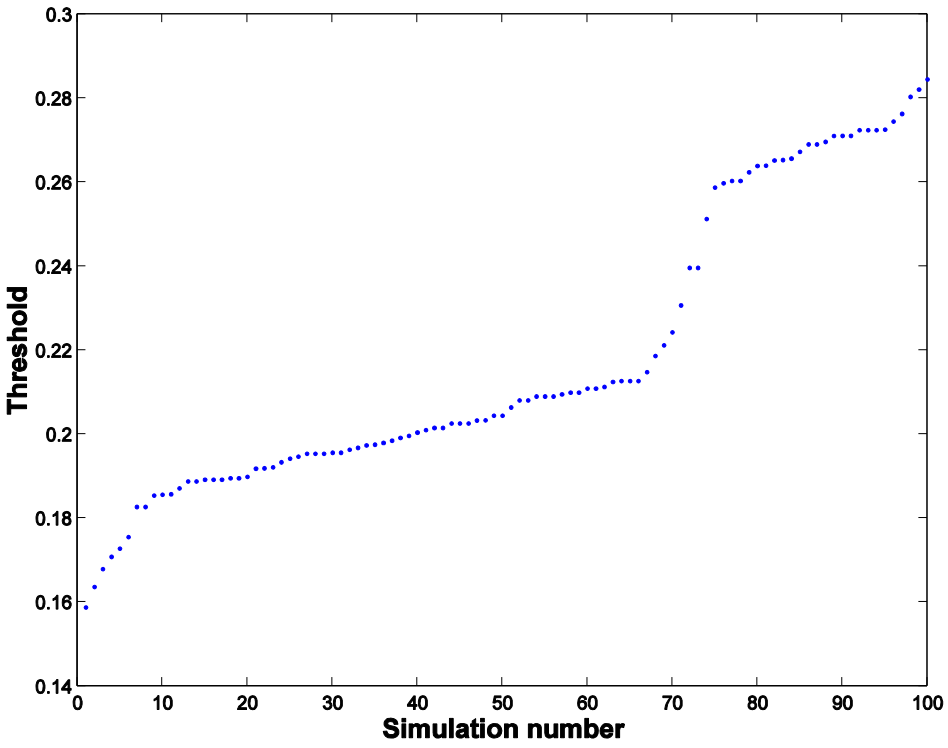


Figure 70: 100 thresholds calculated for independent simulations of method-of-constants experiments, then arranged in order of size.

I originally feared that the level of quantisation would be so great that data might only be drawn from a few distinct threshold levels – perhaps as few as three or four. When I investigated this through simulations, I found that the results from actual datasets were not nearly as poor as I had feared. Although a series of discrete levels are clearly visible, they are numerous enough that this issue is probably not a concern – at least for these particular stimuli. The sigmoid distribution of the data also suggests that the threshold estimates are reasonably normally distributed around the mean estimate.

However, whilst experimenting with the program I wrote to generate figure 70, I chanced upon a particular combination of stimulus levels that seemed to produce a pathological distribution of results. Figure 70 shows this example. Most of the threshold estimates are clustered around two distinct levels, with around two thirds falling into the lower distribution and one third into the upper distribution. Although I still do not completely understand the reasons for obtaining this distribution, the significance of the problem is clear. Unless a reasonably large number of thresholds are drawn from this population, estimates of their variance might well be grossly underestimated, greatly increasing the chances of making a type-I error. If outlier-removal were to be performed, the chances of making a serious error might be increased even further.

The configuration of stimulus levels and probabilities that led to this situation was aberrant in several ways. Performance at most of the stimulus levels was close to either 0 or 100%, and only one stimulus level was truly 'variable'. Yet similar situations may have occurred during some of my first experiments, where the slopes of the psychometric functions were so steep that it was difficult to place the constant stimuli such that performance was neither 0% nor 100%.

One obvious answer to this conundrum would be to extract the 'variance' from the confidence intervals, and feed it to the ANOVA function directly. Whilst this may be relatively straightforward, despite seeking a consultation with a statistician I have not been able to determine how it might be accomplished. It is always difficult to 'prove a negative' – i.e. that such methods do not exist – but neither have I been able to find any discussion of such a technique in any textbook of statistics at my disposal (Lindman 1992; Robson 1994; Hopkins, Hopkins et al. 1996; Cohen 2007); neither did a 'Web of Science' search uncover any pertinent information.

ANOVA in practice

Typically, individual experiments are broken down into a number of sessions in order to avoid observer fatigue. Individual sessions may further be broken down into smaller blocks of just a few trials for each stimulus level. Standard implementations of 'ANOVA' may then be used to calculate 'p' values for the differences between two or more psychometric functions, as the inter-block variances are calculated natively. However, this approach is not mentioned in any of the textbooks cited in the previous paragraph, and although my literature survey discovered several papers that used ANOVA, none of them provided any references or confirmation that this is exactly how the technique might be performed.

This posed several problems for me. As I have only a surface understanding of statistical techniques I am wary about carrying out procedures that are not well documented, or of applying well documented techniques in situations that appear to differ significantly from 'textbook' examples of their usage. I felt that using ANOVA in this way fell into both these categories as I could see several potential pitfalls. Also I did not collect my results with the intention of performing significance-testing on them as I was not then aware that it was a possibility. I made my experimental blocks as long as I considered humane, because shortening the blocks significantly reduces experimental efficiency due to the long periods of adaptation required at the start of each block during which responses are not counted. In many of my later experiments I carried out complete sets of observations for each condition in a single block of trials. If inter-block variances were to be used as a basis for ANOVA then it would be necessary to split these blocks into smaller chunks for analysis.

I therefore decided to perform a brief sanity check to test whether the results of such a technique would be robust to the number of blocks chosen in each experiment, and whether they would be approximately comparable to 'p-values' determined using 'pfcmp' which is limited to comparisons of only two psychometric functions. At present, I consider 'pfcmp' to be my gold standard as it was written specifically to deal with this class of data and is well integrated into the 'psignifit' toolbox.

ANOVA compared with Monte-Carlo tests of simulated datasets

Monte-Carlo simulation has a time honoured tradition in the validation of statistical methods. Student originally verified the T-distribution through an early application of Monte-Carlo methods (Student 1908). More recently, the technique has been applied to the validation of statistical techniques in the face of violations of their assumptions e.g. Boneau (1960). Here I employ a similar method to investigate the validity of applying ANOVA to psychometric data.

After fitting a psychometric curve, the 'psignifit' toolbox returns the generating probabilities for the (assumed) underlying psychometric function. The stimulus intensities, in my case cone-contrasts, are already known from the experiment. This enables us to generate simulated datasets with similar properties to the original data. They can be used to investigate features such as the type-I error rate that might be expected, in an artificial situation where we know that the null hypothesis is true and can be certain of eliminating all forms of bias.

I decided to use simulated datasets of 192 trials at each of 8 points. Each simulated dataset was prepared based on the psychometric function fit to data from one of my experiments by psignifit. The number of trials – 2^5 x 6 = 192 was decided because psignifit is unable to work consistently with datasets of less than six trials per point. I wished to investigate the effect of choosing different block sizes, and by choosing this multiple of a power of 2, up to five different block sizes (2, 4, 8, 16 and 32 trials per block in 32, 16, 8, 4 and 2 blocks respectively) can be investigated.

Testing the type-1 error rate

To test the type-1 error rate, two randomised datasets were generated using identical generating probabilites, then analysed by both ANOVA and pfcmp in order to generate p-values. This was repeated 2000 times, and the proportion of p-values below threshold α was calculated, and 95% confidence intervals for these proportions were calculated using the Matlab function binofit. This was performed for α=0.05, 0.01 and 0.005.

The results are favourable to both methods, showing that each produces type-I errors with the expected frequency, to within the limits of the confidence intervals. This is true for all five block sizes assessed using ANOVA, but it is also true for ‘pfcmp’. Figure 71 shows the mean and standard deviation of p-values obtained by comparing datasets, generated randomly using the same generating function and analysed using ANOVA after splitting the datasets into 2, 4, 8, 16 or 32 blocks. There is no difference in the mean p values obtained in each case, or in their standard deviations. Further data are shown in table 38.

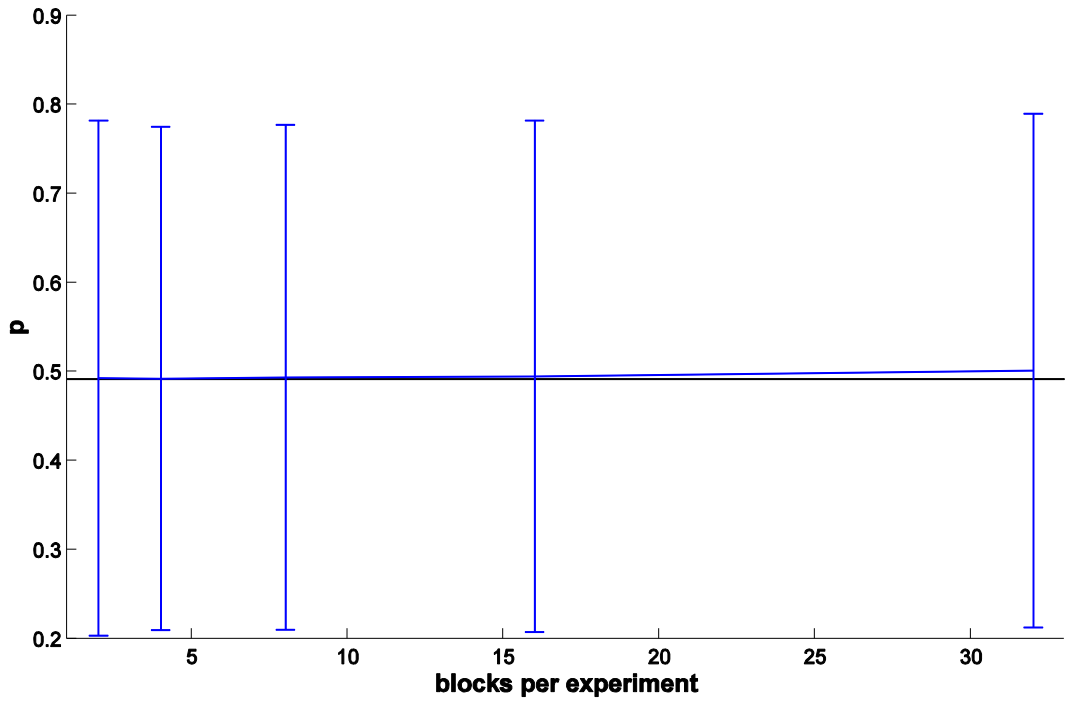


Figure 71: mean and standard-deviation of p-values obtained for randomly generated datasets using identical generating functions.

This pleasing outcome is partly offset by the observation that there is surprisingly little agreement between the results obtained with pfcmp and ANOVA, or indeed between the five block-sizes tested with ANOVA when applied to identical datasets. Although there is clearly a degree of correlation, it is quite possible for one test to return a highly significant p-value but the other to return a value several orders of magnitude higher. Even when only ANOVA is used to analyse the same data, it is common for the p-values obtained to differ considerably for each of the five block sizes.

Number of repeats:
2000
Alpha value: 0.05

		Anova:				
	pfcmp	2 blocks	4 blocks	8 blocks	16 blocks	32 blocks
n (p < alpha)	101	97	96	104	98	110
n / 2000	0.0505	0.0485	0.048	0.052	0.049	0.055
95% lower	0.0413	0.0395	0.0391	0.0427	0.04	0.0454
95% upper	0.061	0.0588	0.0583	0.0627	0.0594	0.0659

Table 38: comparison of p-values obtained for 2000 randomly generated datasets, using different block lengths and methods of analysis. For 2000 repeats, we would expect an average of 100 p-values to be smaller than an alpha cutoff of 0.05. "n/2000" calculates the exact proportion, and 95% upper and lower bounds for this estimate are displayed directly below, calculated using the binomial theorem.

	Overlap:				
Condition	2	4	8	16	32
pfcmp	31.9588	35.4167	31.6832	21.4286	8.9109
2		67.7083	54.6392	40.2062	19.5876
4			63.5417	47.9167	23.9583
8				55.102	25
16					34.6939

Table 39: the degree of overlap congruence for the two results, comparing the method 'pfcmp' with ANOVA performed on the data after they were divided into different block sizes. Each figure is calculated by dividing the number of p-values judged significant by both methods (p<0.05) relative to the mean number of p-values judged significant by either method, then multiplying by 100 to obtain a percentage.

I am uncertain how best to measure correlation in this sample. Two obvious measures are to ask what proportion of the significance tests are in agreement with each other, or to perform a correlation test such as performing Cronbach's alpha (Cronbach 1951). The results of the first approach are shown in table 39: there is a roughly 10-30% overlap between results judged significant using pfcmp or ANOVA at the 5% level. Although these figures are 4-12 times better than would be expected by chance, my feeling is that the finding does not cast the method in a convincing light. Very similar patterns of results are found for alpha values of 0.01 and 0.005.

Cronbach's alpha – which has no direct relation to the level of significance, alpha – uses Spearman's ranking algorithm, and may therefore be used to analyse skewed or normally distributed data. In my simulations it returns values of between 0.65 and 0.69 when used to compare results obtained by pfcmp and ANOVA, depending on the number of blocks. These figures would be considered unacceptably low in other fields, such as the development of psychometric or clinical tools (Bland and Altman 1997). This should seem disconcerting: a test has little value if it returns values that are only loosely based on the data, even if the statistical properties of the results do correspond closely to the predicted distribution.

As a comparison, I performed a similar test on pseudorandomly generated datasets analysed using both ANOVA and the Kruskal-Wallis nonparametric test. I determined p-values for randomly generated datasets of 50, 100 and 500 'trials', normally distributed and with identical variance for two conditions. Congruence between the results of the two methods was usually around 80%, whatever the critical level of significance, and Cronbach's alpha, determined for the resulting p-values, was greater than 0.95.

Tests of datasets known to be different

In these demonstrations, the first dataset was created as before, but it was copied exactly to produce the second dataset. To introduce a difference in the threshold between the two psychometric functions, the contrast levels of the second dataset were translated linearly in order to cause the threshold differences between the two curves to become larger or smaller.

After the two simulated datasets had been prepared, they were copied 5 times, and split into 2, 4, 8, 16, or 32 equal sized blocks. Each block was fit with a psychometric function using psignifit. For each of the 5 pairs of datasets, ANOVA was performed in order to determine whether the two psychometric functions were significantly different from one another. P-values were then plotted against the number of blocks into which each dataset had been divided.

The initial result was that the p-values varied both widely and wildly, and there was no clear pattern in the data. I repeated the ANOVA analyses 128 times, shuffling the order of the trials in the simulated datasets to a new, random order each time, but without changing the proportions of positive/negative responses for each contrast level. The resulting p-values were converted logarithmically then averaged for each block. The resulting pattern was a lot clearer: as the data were subdivided into smaller and smaller blocks, the p-values were reduced – sometimes by several orders of magnitude. Repeating the analyses in this way also enabled me to calculate the standard deviations of the data thereby obtained, which were surprisingly large.

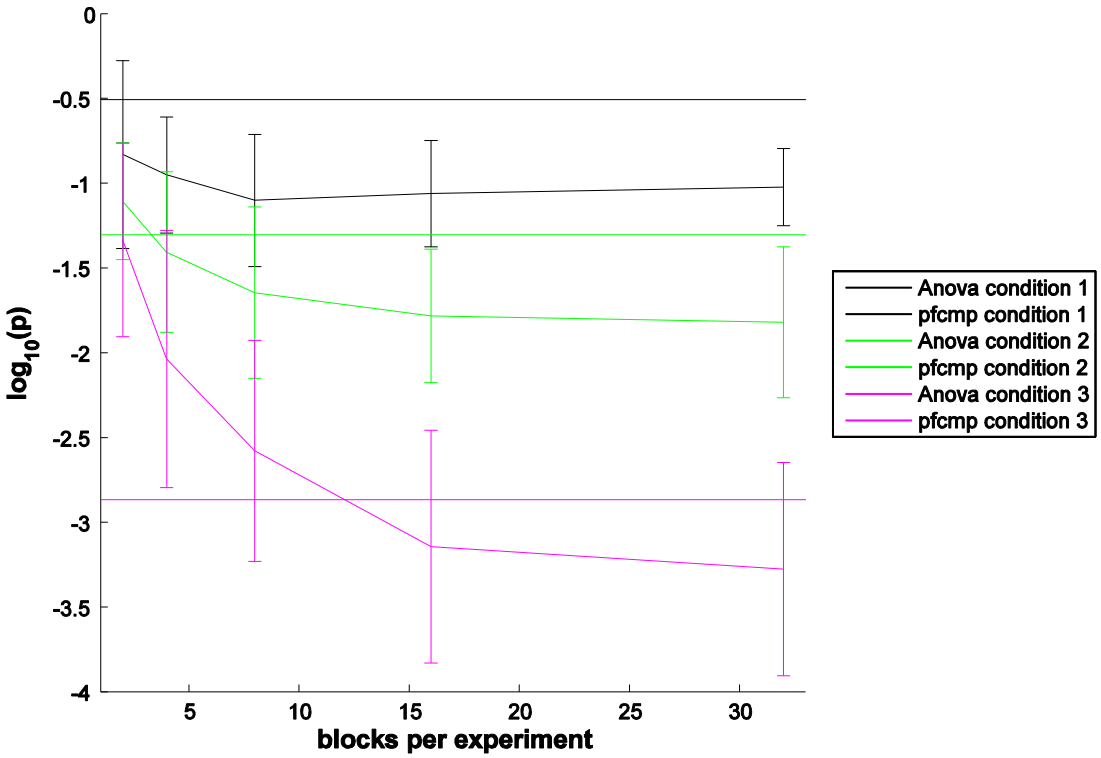


Figure 72: 'p-values' calculated if 2 simulated datasets of 192 observations are split into 2, 4, 8, 16 or 32 blocks of 96, 48, 24, 12 or 6 trials per contrast level, and the threshold values for each block are compared using 'ANOVA'. Error bars show the standard deviation for the estimates of the p-values, based on the 128 repeats. 95% confidence-intervals would be too large to display clearly.

For comparison, p-values generated by running 'pfcmp' are shown, which were generated by running the test on the original datasets of 192 trials per condition. Confidence intervals are not provided for the 'pfcmp' function, as the variance of the output each time it is run is simply a function of the number of simulations chosen. However, even for modest numbers of runs, the standard deviation of the p-values output by 'pfcmp' is an order of magnitude lower than those of 'ANOVA'.

These results, plotted in figure 72, show that using ANOVA to test psychometric functions for significance is not straightforward. Firstly, the results are strongly dependent on the number of blocks into which the experiment is divided. This is important, because the simulated data are that of an ideal observer who exhibits statistical stationarity - who does not fatigue and whose observations are independent of the length of the blocks or their position within the sequence of observations. The number of blocks into which an experiment is subdivided, and rearrangements of the trial order should therefore have no bearing on their statistical significance. Secondly, the standard-deviations of the p-values obtained are often so large that they might often be considered almost random for practical purposes, even though the only difference introduced was to re-order the results of the trials by shuffling. A single ANOVA performed on two curves with an average p-value on the border of statistical significance at the 0.05 level, might give a value anywhere between 0.2 or 0.005.

It is important to emphasize that this phenomenon differs from the normal variation in p-values that is to be expected even if an experiment is repeated, even under identical conditions. This occurs because whenever data is collected afresh, it will be subject to the laws of statistical sampling. However, here we are simply re-analysing the same raw data and an ideal method would therefore provide the same p-value on each occasion.

As a Monte-Carlo method, 'pfcmp' provides slightly different 'p' values each time it is run on identical data. However, the spread of these values is dependent on the number of iterations run in each simulation, which can be weighed against processing time to achieve arbitrarily small degrees of precision. The error of the estimate is also dependent on the absolute value of the calculated p-value. For example, if 10,000 simulations are run they are sufficient to calculate a p-value of 0.05 with error bounds of +/- 0.0056. However, if the actual p-value is 0.001 then the 99% error bounds will be 0.0005 to 0.0022, which is proportionately much greater. For the sake of clarity I chose not to show the standard deviation of the p-values obtained using pfcmp.

Could a simple correction be applied to the data, in order to compensate for the number of blocks into which the experiment was divided? Or is ANOVA valid for one particular number of blocks? figure 72 would appear to rule this out, as the crossover point between the ANOVA results and the pfcmp results is different at different levels of significance. However, it could be that the ANOVA values are valid, but only for large numbers of blocks, where the curves appear to approach their asymptotes. This would actually imply that ANOVA is more powerful than 'pfcmp' when used appropriately, which would be desirable.

One possible resolution would be if both 'pfcmp' and 'ANOVA' provide valid p-values, but that they are asking different questions. 'pfcmp', as previously discussed, takes note of a number of different parameters of the curve – notably the threshold, slope and lapse-rate. ANOVA operates under the assumption that the datasets being compared have identical variances (here, the variance will correlate strongly with the slope) and ignores the lapse rate. Perhaps incorporating these assumptions allows ANOVA to provide smaller p-values than it would otherwise do.

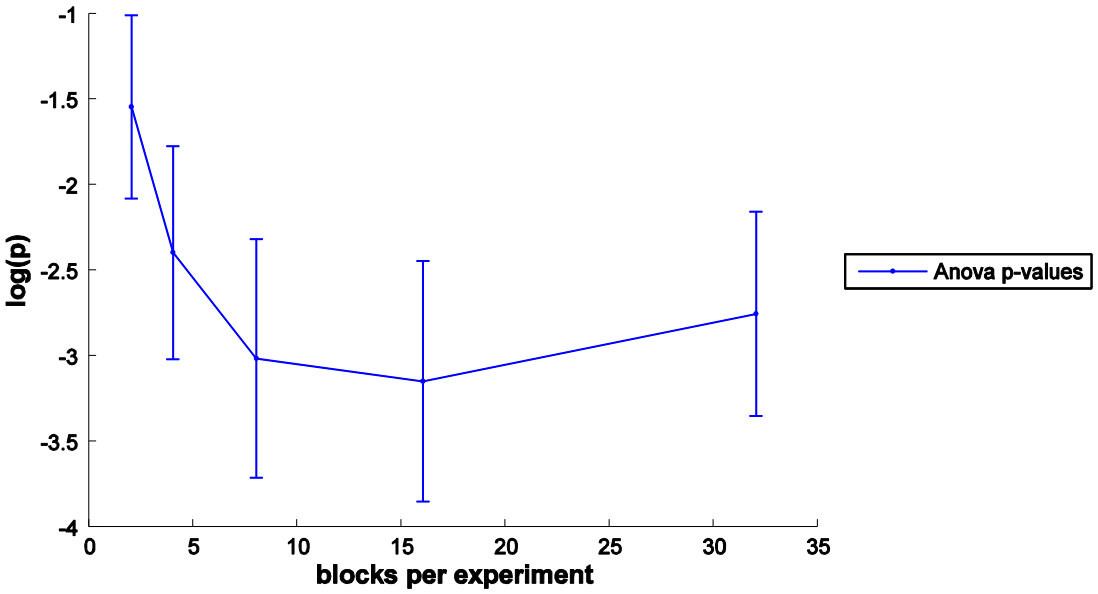


Figure 73: 'p-values' calculated if 3 simulated datasets of 128 observations are split into 2, 4, 8, 16 or 32 blocks of 64, 32, 16, 8 or 4 trials respectively, and the threshold values for each block are compared using 'ANOVA'. Error bars indicate the standard deviations.

Figure 73 further shows the results when ANOVA is applied to three psychometric functions rather than two. The comparable 'pfcmp' p-value is not plotted, as pfcmp is unable to compare more than two psychometric functions directly. As for figure 72, the resulting p-value is strongly dependent on the number of blocks or sessions into which the experiment has been split. Yet the dependency of the p-values obtained on the number of blocks into which the experiment is divided remains a fundamental problem.

A possible objection to the method used here, is that the repeats used the same data, so the samples were not independent as they would be in a real experiment. For example, if a simulated dataset is split into two blocks then if the psychometric curve fit to one block is very close to the psychometric curve that would be fit to the combined dataset, it stands to reason that the psychometric function fit to the second block should also be close to the threshold that will be obtained if the whole dataset is analysed as a single block. The converse is also true. To determine whether this interaction might have caused my ANOVA results to be invalid, I repeated the simulation but re-randomised the dataset for each repeat. The resulting graphs (not shown) showed a nearly identical pattern to the first simulations, save that the error bars were even larger because they were determined both by the increased variance of the underlying data, and by the inherent variability in the technique.

Pooling data from multiple observers

An important objection to the use of ANOVA is that its employment would necessarily involve pooling data from several observers in order to obtain a sufficiently large number of datapoints. There seems to be a practical lower limit of about 6 trials-per-point when subdividing datasets for analysis; for smaller datasets, 'pfit' often fails. Depending on the experiment, individual observers often made only 20-40 trials per point per condition, meaning that only 3-6 thresholds are available per condition per observer. Either of these figures is small in comparison with the numbers of measurements for which t-tests (equivalent to ANOVA performed on 2 groups) are normally used.

A further problem with this approach relates to the manner in which I collected the data. The datasets are different in length for each observer. My own data are typically more extensive than those for other observers such as project students who could only devote a limited amount of time to a project. Dealing with this through ANOVA would either require some data to be discarded in order to equalise the number of blocks per-observer, or for an unbalanced design of ANOVA to be employed. Data would also have to be discarded if the number of trials per point was not a figure divisible by six – or whatever number of trials-per-block was chosen. Finally, sometimes the numbers of trials-per-point-per-block differ because of issues such as observer-sessions being cut short by fire alarms. In these situations, data was not discarded and re-collected. This did not matter for the algorithms that calculate confidence-intervals, but is clearly significant for ANOVA employed this way.

My real objection to performing a pooled ANOVA is that it is not cost-free. Calculating a p-value in this way implies an acceptance that it is necessary to do so. It then becomes unseemly to display or to comment on individualised data if a p-value cannot be calculated for them. If it is possible to interpret data without pooling it between observers, then it is my view that it is generally preferable to do so.

Summary

ANOVA *may* be a valid method of calculating p-values when sufficient numbers of blocks are available. However, my statistical knowledge is insufficient to be able to *show* that this is so. For datasets that may be split into sufficiently large numbers of blocks, my simulations suggest that ANOVA is probably consistent: the reshuffling error is sometimes small relative to the sampling error if enough blocks can be sampled. But this 'magic number' of blocks is clearly not a constant, being strongly dependent on the level of statistical significance, the variance of the data and the length of the blocks. The fundamental question that I have not been able to answer is: "how I can *know* whether I have sufficient numbers of blocks for an ANOVA test to be valid?" My overwhelming feeling is that in the absence of additional guidance, performing ANOVA on inter-session variances is a fundamentally poor choice of technique.

Logistic regression

Logistic regression is a statistical technique closely related to ANOVA; both are 'generalised linear models'. However unlike ANOVA it is able to deal directly with binomial or multivariate (categorical) data. Following a statistical consultation, I attempted to use binomial logistic regression (BLR) to analyse my data. However I ran into problems caused by the fact that many – perhaps most – of the psychometric data that I wished to compare were not sampled at identical stimulus contrasts. If I ran BLR on such data, both Minitab and SPSS flagged errors and refused to continue.

An obvious solution would be to ensure that psychometric functions to be compared were always sampled at the same levels of stimulus contrast or intensity. However, this requirement would often involve collecting large amounts of data at inappropriate stimulus intensities, where the observer would always make the same response. This would be fatiguing for the observer, and might increase the chance of observer bias as positive and negative responses could no longer be balanced – even approximately – for each condition. Finally, this solution would come too late for me, as I had already collected my data when I was asked to perform additional analyses.

To deal with this problem post-hoc I was advised to bin the data according to stimulus contrast, using a small number of bins. Unfortunately although this is may be an acceptable method in many circumstances, it can cause severe problems with psychometric functions that are already only sampled at a relatively small number of distinct stimulus intensities. I have prepared some brief demonstrations of why this is so.

Binning artifacts

The two following artificial scenarios demonstrate possible pitfalls of binning data prior to running logistic regression. They are based on simulated datasets, modeled on a cumulative Gaussian psychometric function that may be linearly translated along the axis of stimulus-intensity to create psychometric functions with identical slopes, hits and misses, but with different thresholds. These are then analysed using both 'pfcmp' and logistic regression.

For clarity, the psychometric functions in the first two examples are only sampled at four points and represent data that might be obtained from a single-interval experiment. The two separate conditions being tested are represented by red and blue bars respectively, and the underlying psychometric functions are overlaid on top of the original data. 512 simulated trials were generated at each point.

Scenario 1

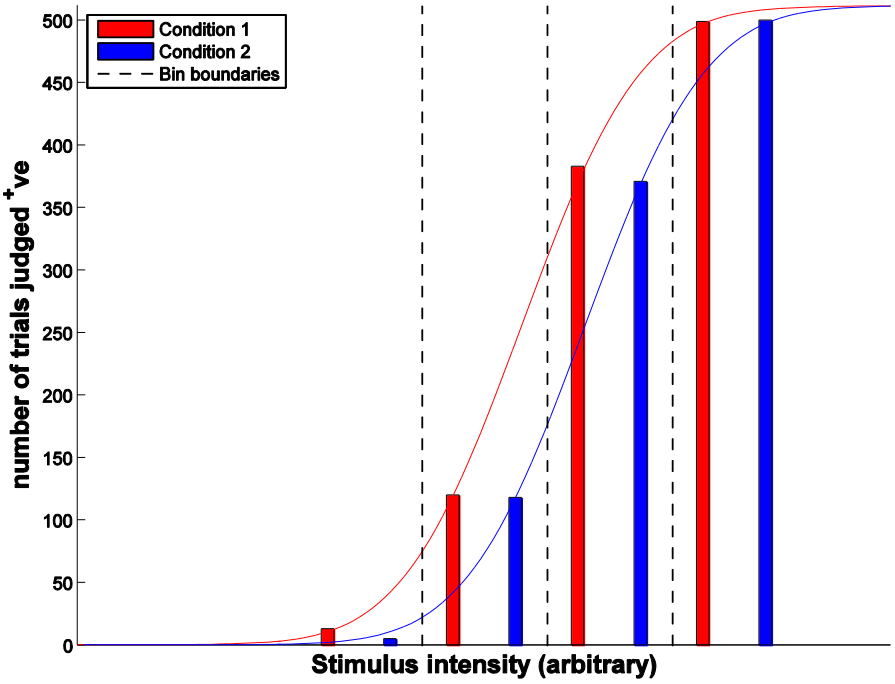


Figure 74: Two different psychometric functions sampled at different values of stimulus intensity. The vertical, hashed lines indicate the boundaries of the four bins. The sampling intensities were adjusted in each case so that they sampled the underlying psychometric functions in approximately the same manner. This is a procedure generally employed by psychophysicists in order to collect data as efficiently, as there is little point in collecting responses for stimulus intensities which will be judged entirely positive or negative. This procedure results in proportions of positive responses in each bin – indicated by the height of the bars – which are roughly equal for both psychometric functions.

In the first scenario, illustrated in figure 74, the two underlying psychometric functions differ. However, because the sampling intensities were adjusted to match the functions the data become near-identical after they have been placed into four bins so that they can be analysed using logistic regression ($p = 0.975$). However, 'pfcmp' correctly determines that the data were produced by different psychometric functions ($p = 0.000$).

Scenario 2

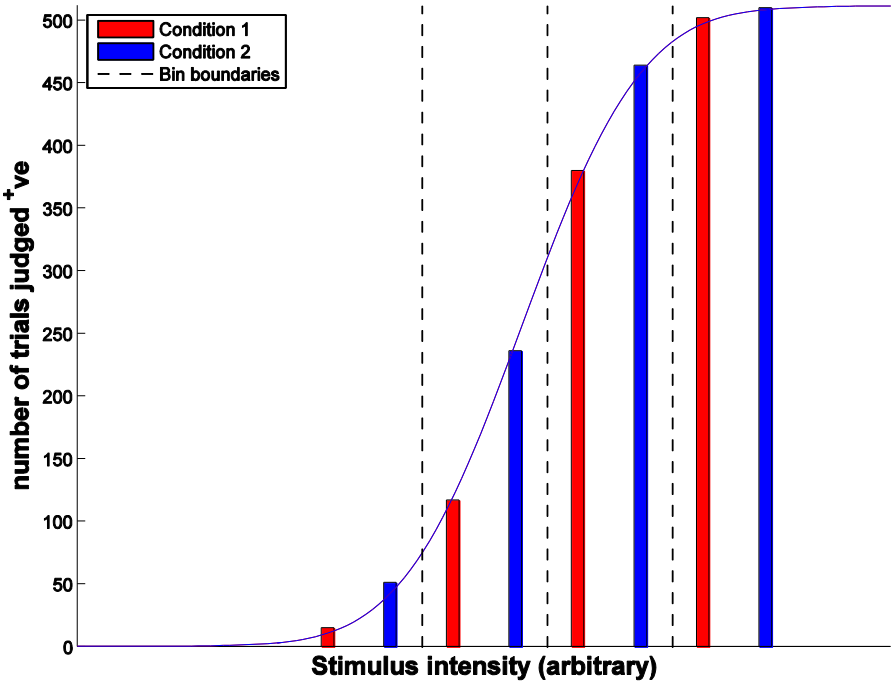


Figure 75: Two identical psychometric functions sampled at different degrees of stimulus intensity. This situation contrasts with that of figure 74, because although the sampling intensities were identical in both cases, in this instance different numbers of +ve stimulus intensities are obtained for the two psychometric functions.

In the second example illustrated by figure 75, the psychometric functions are identical but have been sampled at different stimulus intensities, leading to different numbers of positive responses for each condition. After binning and analysis using logistic regression, the two conditions appear to be different with a high degree of confidence ($p = 0.000$). Conversely, 'pfcmp' correctly finds that the two conditions are not statistically significantly different ($p=0.969$).

In summary, these examples are highly artificial, but serve to highlight some of the problems that would occur if logistic regression were used to analyse real psychometric data.

Conclusion

Of the four methods discussed in this chapter, only 'pfcmp' and the visual comparison of error bars appear to have any degree of validity. Whilst other forms of ANOVA or binomial logistic regression may be valid, I have been unable to find information about how to employ them correctly.

Appendix F - Applying NHST to a Problem in Psychophysics

Introduction

In this appendix, I am going to walk through an analysis of data similar to that from one of my experiments, illustrating some of the practical problems, and the pitfalls to the interpretation of the results.

Multiple hypotheses supported by multiple tests

The data from figure 77 can also be displayed in the form of several overlapping ovals. They originate in experiment 2.2.4, where I provide evidence that simultaneous chromatic induction is tightly tuned along the cardinal opponent axes. I show graphs for each of three observers, and on each graph there are three separate curves of datapoints. Two show the results of induction on the cardinal axes; one shows the results for induction along an intermediate axis. For the two cardinal-axis induction curves, induction is generally weak, but it is strong for particular textures. This was predicted by previous experiments, but replicated again for each observer in these experiments.

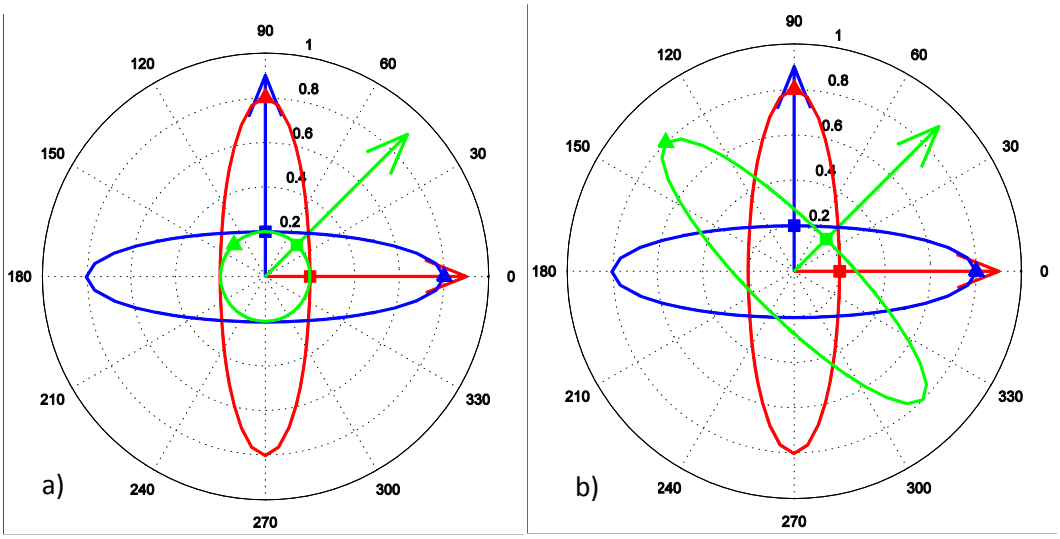


Figure 76a and b: A) shows the pattern of results that I predicted; B) shows the pattern of results that other researchers generally told me that they would have predicted.

The blue and red axes represent induction for background shifts along the cardinal chromatic axes. The green ellipse represents induction along an intermediate axis. The radial angle represents the chromatic texture within the target square. For a deeper explanation of this diagram see the discussion of experiment 2.2.4.

What I *intended* to investigate in this experiment, is whether or not the intermediate axis is equivalent to the cardinal axes, or whether it is in some way qualitatively different – in which case I was moderately sure that induction would be roughly equally weak for all the textures. When I undertook the experiment I genuinely did not know what the result would be. If pressed, I would have predicted the actual outcome (that the cardinal and intermediate axes would differ) – shown in cartoon in figure 76a. But from conversations at conferences, my impression was that most researchers in the field would have predicted the opposite finding. They expected the intermediate axis to behave similarly to the cardinal axes. There would be a single texture for which the intermediate-axis condition showed strong induction, and the axis of this texture would be orthogonal to the axis of induction – this is the pattern shown in figure 76b.

NHST involves, in formulations of which I'm aware, at most two hypotheses: a null and an alternate. Furthermore, it is considered to be vital to choose the hypothesis to be tested prior to starting the experiment rather than collecting the data then choosing the hypotheses post-hoc. Even if I were to 'cheat' and choose a null and alternate hypothesis having already decided what pattern I *think* the data shows, should I choose the null-hypothesis whose refutation supports my theory, or should I choose the null-hypothesis whose refutation seems to support the theory that most other researchers seem to believe in? Or should I perform both analyses, both refuting the generally accepted theory and failing to refute my own?

The hypotheses to be tested are that for the intermediate axis, there will be no texture setting for which induction is strong, and that the resulting ovals will not be strongly oriented to any particular angle, as the cardinal axes are. However, justifying each of these findings may require several different hypothesis tests. Examining them each in turn, the difficulties in testing them become clear:

1. What do I mean by ‘strong’? In this situation, I am referring to induction along the cardinal axes, which as discussed is relatively strong for certain textures. But how strong would induction need to be before I rejected the hypothesis? Using the Neyman-Pearson model of NHST, it would be possible to perform a one-tailed test to see whether induction was stronger in the intermediate axis than in either of the cardinal induction conditions, but this would be too stringent a test. If induction in the intermediate axis was nearly but not quite as strong as it is in the cardinal axes, we would falsely accept the null-hypothesis. If we performed 2-tailed tests, we might reject the null-hypothesis that induction was equally strong for the intermediate axis and either of the cardinal axes, and make the same mistake. Clearly the predictive theory needs to be made more precise, but I am uncertain how to reformulate it. The only way I can think of is to predefine an arbitrary threshold at which I would start to accept that induction in the intermediate axis was ‘large’.

2. The second hypothesis is equally fraught. How is a non-directional oval defined? The answer is that it is a circle. If induction for all the textures were non-significantly different for this axis, then we could accept that we had failed to prove non-circularity. Succeeding in proving that all the cardinal-axis ovals were oval and not circular would provide some evidence to suggest that the test was powerful enough to differentiate between induction in the cardinal axes and induction in the intermediate axis. But what if we increased the power of the test by collecting more data? It is inconceivable that the null-hypothesis (that induction is equally weak in all textures) is actually true. Ultimately we would have to look at the size of the effect, and whether it is much smaller for the intermediate axes than for the cardinal axes.

The issue here is that by multiplying the number of hypotheses that must be considered, it becomes more likely that at least one of them will be confirmed. If we agree to test multiple hypotheses from this data, then we should correct the resulting p-values to compensate.

Contrariwise, demonstrating that each line plotted on the graph shows what I argue it does involves carrying out a fairly large number of individual hypothesis tests. Although it would be possible to state whether or not the intermediate axis is directional without reference to the cardinal axes, if the cardinal axes showed marginal or zero tuning this would cast a negative result for the intermediate axis in a less-than-convincing light. Here, the risk of making at least one error increases dramatically due to the number of comparisons that must be made – at least six and perhaps as many as twelve, depending on how many assumptions one is willing to leave untested. How should I combine the corrections in order to reach a single conclusion? I have been unable to find a solution, yet if I cannot answer this question numerically, what rationale is there in performing quantitative statistical tests as a basis for my conclusions?

Testing hypotheses, or finding patterns?

An additional problem is that the data are rich enough to find patterns that have not been predicted. Each curve is sampled at seven different points, and these may be illustrated as in figure 77. An assumption not yet addressed is that the seven datapoints do indeed form an oval or circle. Yet in order to assess 'ovalness' we have been throwing out five of the datapoints and only comparing those at the cardinal axes of our putative ovals.

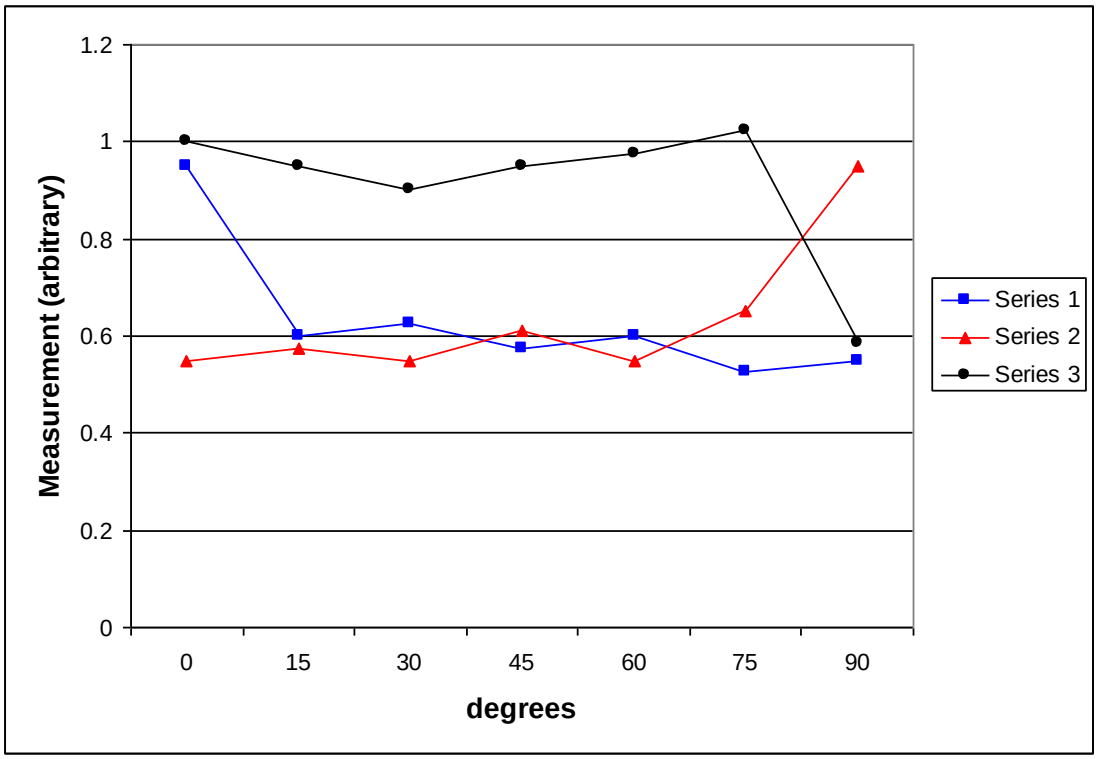


Figure 77: an example loosely based on data from experiment 2. Series 1 and 2 only differ from one another at the endpoints, and conform to a pattern predicted in another experiment. Series 3 does not differ from series 1 at its endpoints, however it clearly follows a different pattern.

Figure 77 shows a graph of imaginary data loosely based on experiment 2.2.4. It is simply an alternative way of plotting the ovals shown in figure 76. Series 1 and 2 conform closely to the oval patterns predicted by a previous experiment, where six out of seven measurements would not be expected to be very different from one another, but one of the endpoints would (either the first or last endpoint respectively). Series 3 is a purely hypothetical pattern that I could conceivably have obtained from the experiment, but it was created purely for the purposes of the following discussion. If plotted on figure 76, its shape would approximate that of an hourglass.

Are series 1 and series 2 significantly different from one another? If one does not take into account the fact that both conform closely to a predicted pattern, it would be easy to dismiss the two endpoints as outliers and conclude that the two series do not differ. This might also be the case even if all the datapoints were analysed statistically. However, if one excludes all of the data save the endpoints from the analysis we would not be able to detect that series 3, which does not conform to the predicted pattern, is significantly different from series 1. This is obvious visually, but if we make the decision on whether to include the intermediate points based on a post-hoc visual analysis, we introduce a circularity by which the result of our apparently objective statistical analysis ultimately depends as much on our eyeball analysis of the data as on the statistical tests applied to it.

Possible solutions

A Bayesian method of analysis, incorporating priors of the appropriate modeling function, might offer a more appropriate means of statistically evaluating such patterns of results and nullify any objections based on 'Feynman's conjecture' which states that the appropriate modeling function must be chosen prior to the experiment (Gigerenzer 2004). Alternatively, Akaike's information criterion offers a quantitative means of estimating which of a collection of modeling functions fits the data most closely (Akaike 1974). However, I am unaware of a formulation able to accept method-of-constants thresholds as an input, and the problem of choosing the candidate modeling functions remains.

The problem is also related, though not entirely analogous, to that of 'unplanned comparisons'. It differs because this research is too exploratory for there to be any truly 'planned comparisons', and it is perhaps also moderately unusual because of the continuous circular axis (chromatic axis) rather than the categorical scale that it uses (termed 'groups' in most ANOVA implementations). However, the solution is identical in both cases: the results should be replicated in order to determine whether or not the findings are robust. Statistical methods designed for focused investigations of known relationships are inappropriate in this situation.

Appendix G - Checking Stimulus Timing Over Short Durations

Introduction

For visual stimuli that demand very precise control over timing, it is generally necessary to control the display on a frame-by-frame basis. Debugging such experiments can be challenging, as the stimuli are shown too fast for human observers to follow by eye. Slowing the experiment to a crawl allows us to check that a program shows the right combinations of colours or bitmaps in the correct order, but without recourse to further equipment it is impossible to be sure that each trial stage is displayed for exactly the desired duration.

Building a phototransistor based stimulus monitor

To help verify my experiment timings, I decided to build a lightmeter using a phototransistor with a fast response time. I did not have a digital storage oscilloscope on hand, so I made a device that could connect to a computer sound card. Its design is not sophisticated, but it costs only a few pounds to build and is convenient to use. It may prove useful in a number of lab-tasks, including:

- Verifying stimulus timings
- Checking stimuli for transient glitches
- Making rough estimates of phosphor decay times
- Checking that display devices operate at the frame-rate the computer assumes they are using. This is often not a valid assumption for newer display devices, such as most DLP projectors.

Of course, the first two tasks may generally be accomplished by software, but the ability to verify the stimulus timings independently of the stimulus generation computer is reassuring, and allows bugs in the program to be detected.

Data acquisition using soundcards

Sound cards are potentially very convenient data-acquisition devices due to their low price, ubiquity and sample rates in the order of 50 kHz or better. Unfortunately they include input filters that are generally undocumented, rendering the cards generally unsuitable for subversion into oscilloscopes. Furthermore because the inputs are capacitatively coupled they can be used only for AC measurements. The device I describe here allows measurements of the timing and relative brightness of a scanline as it traces across the computer screen, and this is sufficient for many purposes.

Figure 78: the completed timer

This project, which should be very simple, is complicated by the fact that soundcard microphone inputs seem to be very poorly standardised; indeed each of the four soundcards I tested was constructed differently, as described in table 40, even though each takes the same 3.5mm jack inputs.

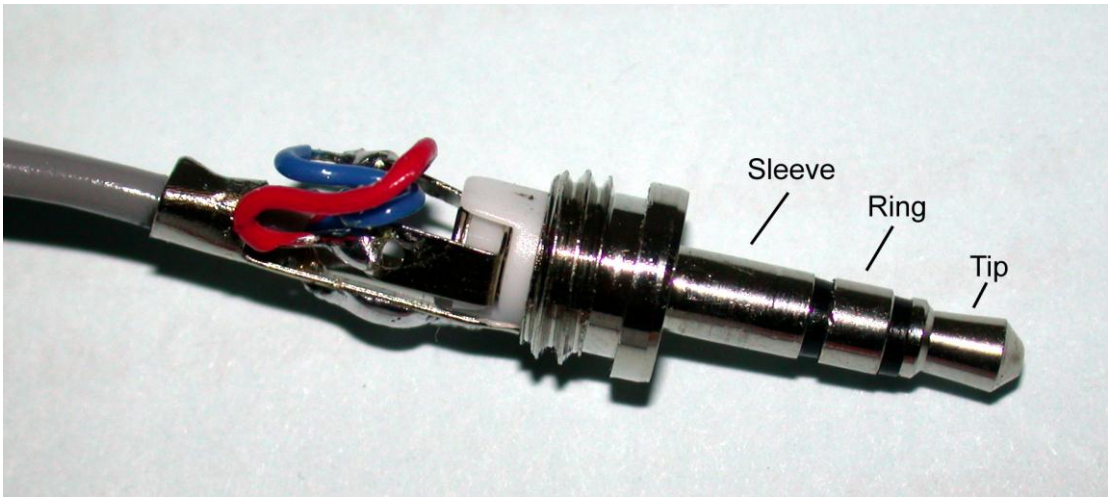


Figure 79: a picture of a 3.5mm jack plug

	Computer			
Jack connections	***PC-1***	***PC-2***	***Dell laptop (? stereo)***	***IBM laptop***
Tip	+5v	+5v	+5v HI	Signal
Ring	Signal (high sensitivity)	Signal (normal sens.)	+5v HI	+5v
Sleeve	Ground	Ground	Ground	Ground

Table 40: Connections to computer stereo microphone jacks vary widely

These inputs are designed to work with cheap electret microphones, which require a 5v bias and may have either two or three pins. Both types of microphone work essentially on the same principle and contain a small FET transistor, whose resistance varies as sound waves strike the microphone. This transistor is connected in series with a fixed value resistor to form a potential divider so as the microphone's resistance changes, the voltage across the transistor does also. In the two-pin devices, the fixed resistor is on the sound card, which may therefore use a single connection to bias the microphone and read the signal. In the three pin devices, the resistor is included in the microphone assembly, and the signal across the potential divider is read from the extra pin. My

phototransistor circuit is identical in principle, except that the FET transistor in the microphone is replaced with a phototransistor whose resistance is modulated by light rather than sound.

Circuit explanation

Figure 80 shows the circuit design for the timer. Certain compromises, for example in the values of the fixed resistors and the omission of a decoupling capacitor for the bias voltage, have been made to ensure that the circuit shown here works on the majority of soundcards and is unlikely to damage them whatever their configuration.

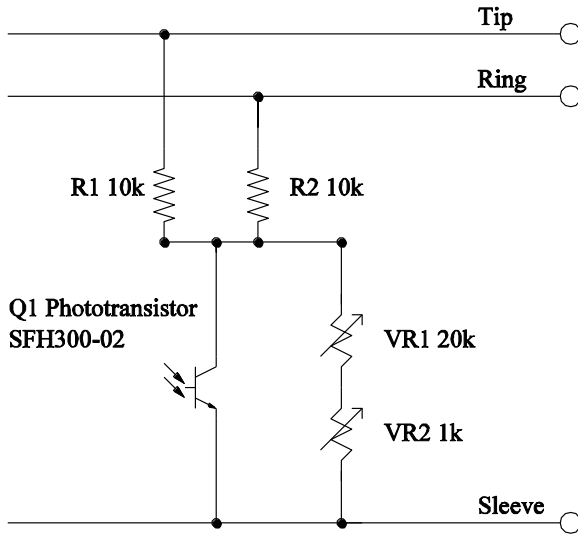


Figure 80: Light meter circuit diagram

If the tip or ring of the stereo jack provides a 5v bias, either R1 or R2 or both may act as the fixed-value resistor in the potential divider. If the stereo jack 'expects' a separate signal-out line, then the extra resistor merely increases the impedance of the output. The phototransistor produces a much stronger signal than an electret microphone would, so the added impedance does not matter, and for high-sensitivity soundcards it may even be a disadvantage. Preset resistors VR1 and VR2 are wired in parallel with the phototransistor and their values may be decreased to attenuate the signal.

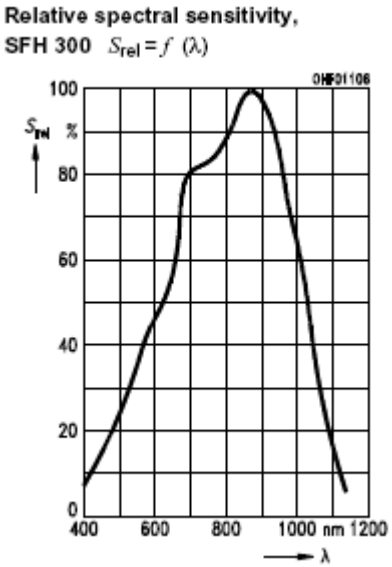


Figure 81: The spectral sensitivity of the SFH-300 phototransistor

Phototransistors are generally either 2 or 3 pin devices and their package may include a filter to reduce their sensitivity to visible light. This circuit is likely to work with most 2-pin phototransistor that have a clear packaging (i.e. without the infra-red filter). The phototransistor used in my device was the SFH 300. Figure 81 shows the spectral sensitivity curve for this phototransistor, taken from the datasheet (Anon 2005). The peak sensitivity is in the infra-red region of the spectrum, but the device is still reasonably sensitive to all visible wavelengths. When assessing phosphor decay times, the decay of each individual phosphor should be measured independently to ensure that the time measured is not biased towards that of the red phosphor.

Figure 82: The completed circuit, mounted in a small box. A suction cup is used to fix the timer to the screen of a display device.

Figure 82 shows a completed circuit mounted in a small project box. The photodiode is directed towards the screen, through a hole in the base of the box. A suction cup is used to fix the box to the screen. A slip of white paper describes for posterity the circuit, the purpose of the device and how its sensitivity may be adjusted.

Using the lightmeter

The lightmeter is held in position against the computer screen by a suction cup, and recordings made by a standard sound recorder - on many linux systems the following command is used:

'# rec –r 44100 myfile.wav'

If computing power permits and provided the machine is able to multi-task, the sound recorder may be run concurrently with the experiment on the same machine. However I preferred to run the experiment independently, using a laptop for data acquisition.

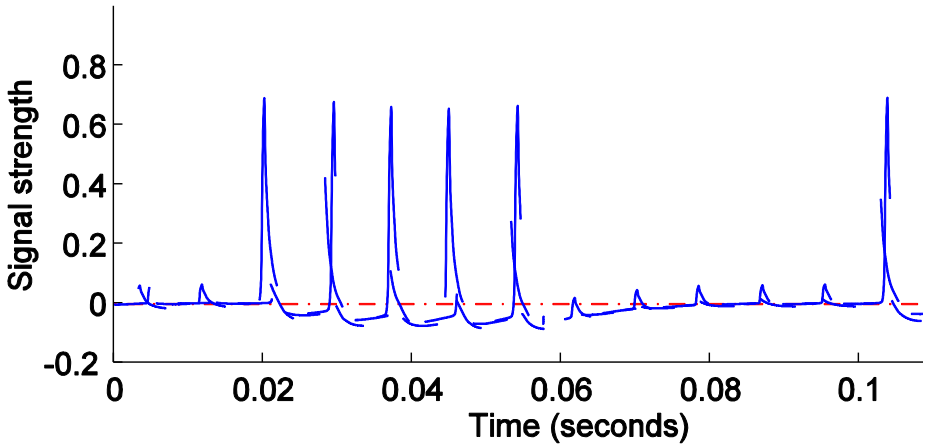
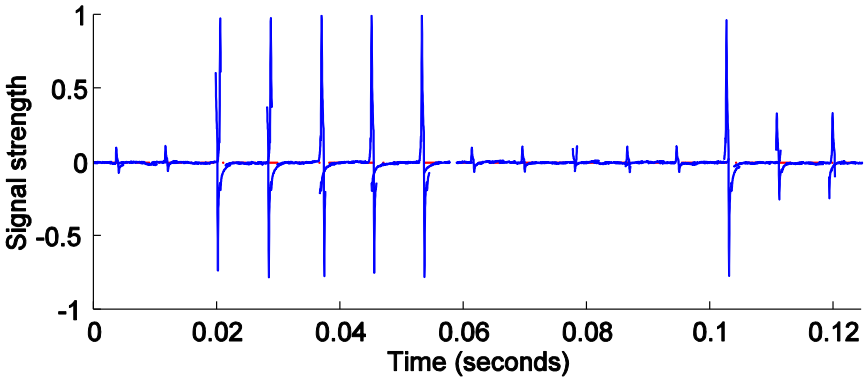


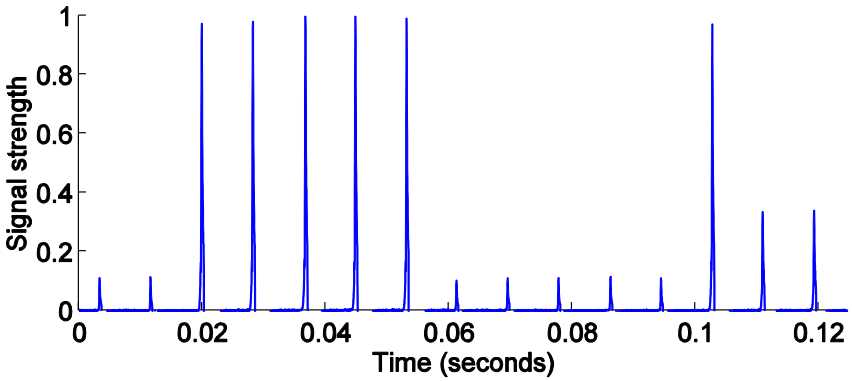
Figure 83: The trace recorded through the soundcard whilst the meter is held close to a CRT monitor

The trace shown in figure 83 was recorded whilst the meter was held close to a standard CRT monitor running at about 80 Hz. At each peak, the meter registered the sweep of the CRT raster past the phototransistor. When the monitor became more or less bright, the amplitude of the signal from the phototransistor became stronger or weaker, though the relationship between peak height and phosphor brightness is not linear. For reference, the first two peaks are small, as the screen was black (RGB = 0, 0, 0). The next five peaks correspond to maximum brightness (RGB = 255, 255, 255) and the final two peaks are of intermediate size (RGB = 85, 85, 85).

The rise time of the phosphor output is much faster than the fall time, illustrating the phenomenon of phosphor persistence. Unfortunately it is not possible to determine the phosphor half-time directly from this graph, due to the non-linearity of the device. However a reasonably accurate reading could be determined if the luminances of some of the peaks were known, were it not for an additional problem: the baseline of the trace varies due to the AC-coupling of the sound card, causing the heights of the peaks to vary depending on their density.



a)



b)

Figure 84: Post-processing of the signal

a) Differentiating the (inverted) signal solves the problem of a drifting baseline. However, each peak is now associated with a trough of a similar magnitude.

b) In this simplified trace, the negative spikes on the graph have been clipped. Three signal levels are apparent, corresponding to periods when the monitor displays white, black or gray frames.

Figure 84 shows that this may be remedied by differentiating the signal from the screentimer. Alternatively the stimulus train could be filtered by a high-pass filter, to remove the DC component. But now much of the data is redundant, as each peak is accompanied by a trough of similar magnitude. To make the plot easier to read, it may be clipped to remove redundant information.

In my isoluminant experiments, there might not be a large change in signal amplitude between different parts of each trial. For testing purposes, each experimental program draws coloured tiles in the center of the screen. For example, the following code optionally displays a strobe preceding each trial so that its onset in the soundcard recording is easy to find.

```
#ifdef debug

if (trial_stage == 1) {for (i=1;i<8; i++) {

  glClearColor((float)(i%2),(float)(i%2), (float)(i%2), 1);

  glClear( GL_COLOR_BUFFER_BIT );

  wait_x_frames( 5 );                    }

                    }

#endif
```

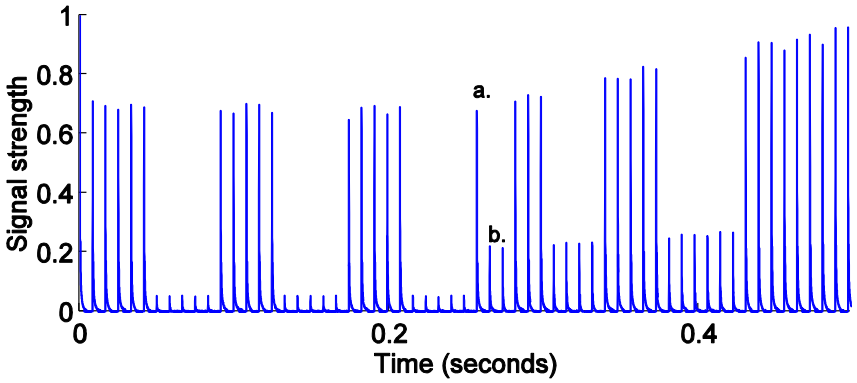


Figure 85: A trace used to verify a test stimulus

A strobe (white – black – white – black – white - black) precedes the onset of the trial, in which a first picture is shown for one frame(a); the second for two frames(b), and so on until the end of the trial (7th picture; waiting for a response so arbitrary duration).

Most of my experimental programs treat each trial as a series of separate pictures, drawn and presented in sequence. Another portion of code appended to the end of the 'draw' function pastes a small coloured square in the center of the display – alternating between a white square for even-numbered pictured and gray for odd-numbered pictures. It takes only a tiny fraction of a frame-duration to draw this additional picture element, but if it did affect the stimulus timings in any way, it would be by prolonging the draw function so that it took longer than a single frame to complete. If the stimulus timings are accurate with the patch in place, then it is likely that they will remain accurate when it is not.

The trace shown in figure 85 precedes the onset of the pre-trial strobe (highlighted in blue) by roughly 0.2 seconds. The trial – starting approximately 0.5 seconds into the recording – consisted of seven pictures shown for respective durations of 1, 2, 3, 4, 5, 6 and an indeterminate number of frames (after displaying the last picture, the program waits for an observer response). It is interesting but somewhat alarming that the brightness of the monitor appears to increase towards the end of this stimulus. I have not yet investigated whether this is an artifact of the monitor or of the screentimer device, filtering in the soundcard – or perhaps even an automatic gain-control mechanism in the microphone input. The spikes appear to increase in size in the undifferentiated plot of this stimulus train, so the problem does not appear to lie in the post-processing.

Having checked the program was working correctly, I wrote some additional code so the computer could monitor its own performance. If timing errors of more than 10% are detected (or suspected) the computer can discount any data from that trial, and repeat it later in the experiment if required. In practice, faults are rare, but the computer appears to be somewhat conservative in reporting timing problems, compared to the screentimer output.

The self-checking method mentioned above was not used for any experiments reported in this thesis. However, in checks of continuous experiments, such as the nulling paradigms, the screentimer was used to check whether the display was updated for each frame, or whether 'glitches' caused certain frames to be repeated for a short time whilst the computer performed tasks unrelated to the experiment. Alternate frames were coloured either white or gray, and a short Matlab script was used to check whether high and low peaks alternated, or were sometimes repeated indicating a glitch

Bibliography

Adelson, E. H. (1993). "Perceptual organization and the judgement of brightness." Science **262**: 2042--2044.

Akaike, H. (1974). "A new look at the statistical model identification." IEEE Transactions on Automatic Control **19**(6): 716–723.

Allan, A. L. (1997). Theory of errors and quality control. Practical surveying and computations. Oxford, Laxton's.

Allard, R. and J. Faubert (2007). "Double dissociation between first- and second-order processing." Vision Research **47**(9): 1129-1141.

Allen, G. (1879). The colour-sense: its origin and development. London, Trübner & Co.

Amano, K., D. H. Foster, et al. (2002). Relational colour constancy across different depth planes. ECVP, Glasgow.

Anon (2005). SFH 300, SFH 300 FA. Regensburg, Osram Opto Semiconductors GMBH.

Anon (2006). ViSaGe Visual Stimulus Generator.

Anon (2007). Translations of CIE 1931 resolutions on colorimetry. Colorimetry - understanding the CIE system. J. Schánda, Wiley-Interscience**:** 1-8.

Anstis, S. and P. Cavanagh (1983). A minimum-motion method for judging equiluminance. Colour Vision: Physiology and Psychophysics. J. Mollon and L. T. Sharpe. London, Academic press**:** 155-156.

APA (1994). Publication manual of the American Psychological Association. Washington, American Psychological Association (APA).

APA (2001). Publication manual of the American Psychological Association. Washington, DC.

Arend, L. E., A. Reeves, et al. (1991). "Simultaneous color constancy: papers with diverse Munsell values." Journal of the Optical Society of America A **8**(4): 661-672.

Ashida, H., A. Lingnau, et al. (2007). "fMRI adaptation reveals separate mechanisms for first-order and second-order motion." Journal of Neurophysiology **97**(2): 1319-1325.

Ashida, H., A. E. Seiffert, et al. (2001). "Inefficient visual search for second-order motion." Journal of the Optical Society of America a-Optics Image Science and Vision **18**(9): 2255-2266.

Baker, C. L. (1999). "Central neural mechanisms for detecting second-order motion." Current Opinion in Neurobiology **9**(4): 461-466.

Barnes, C. S. and S. K. Shevell (2002). "Simultaneous S-cone contrast." Vision Research **42**(1): 75-88.

Barnes, C. S., J. P. Wei, et al. (1999). "Chromatic induction with remote chromatic contrast varied in magnitude, spatial frequency, and chromaticity." Vision Research **39**(21): 3561-3574.

Beauchaine, T. P. (2007). "A brief taxometrics primer." Journal of Clinical Child and Adolescent Psychology **36**(4): 654-676.

Belia, S., F. Fidler, et al. (2005). "Researchers Misunderstand Confidence Intervals and Standard Error Bars." Psychological methods **10**(4): 389-396.

Bender, D. B. and M. Youakim (2001). "Effect of attentive fixation in macaque thalamus and cortex." Journal of Neurophysiology **85**(1): 219-234.

Berger, R. L. and J. C. Hsu (1996). "Bioequivalence trials, intersection-union tests and equivalence confidence sets." Statistical Science **11**(4): 283-302.

Berry, M. J., I. H. Brivanlou, et al. (1999). "Anticipation of moving stimuli by the retina." Nature **398**(6725): 334-338.

Berson, D. M., F. A. Dunn, et al. (2002). "Phototransduction by retinal ganglion cells that set the circadian clock." Science **295**(5557): 1070-1073.

Bland, J. M. and D. G. Altman (1997). "Cronbach's alpha." British Medical Journal **314**(7080): 572-572.

Bleicher, S. (2004). Contemporary Color Theory and Use, Cengage Learning.

Bloj, M., D. Kersten, et al. (1999). "Perception of three-dimensional shape influences colour perception through mutual illumination." Nature **402**: 877-879.

Boneau, C. A. (1960). "The effects of violations of assumptions underlying the t-test." Psychological Bulletin **57**: 49-64.

Bouvier, S. E. and S. A. Engel (2006). "Behavioral deficits and cortical damage loci in cerebral achromatopsia (vol 16, pg 183, 2006)." Cerebral Cortex **16**(10): 1529-1529.

Bowmaker, J., J. Parry, et al. (2003). The arrangement of L and M cones in human and a primate retina. Normal and Defective Colour Vision. J. D. Mollon, J. Pokorny and K. Knoblauch. Oxford, Oxford University Press**:** 39-50.

Bowmaker, J. K., S. Astell, et al. (1991). "Photosensitive and photostable pigments in the retinae of Old World monkeys." Journal of experimental biology **156**: 1-19.

Box, G. E. P. and S. L. Andersen (1955). "Permutation Theory in the Derivation of Robust Criteria and the Study of Departures from Assumption." Journal of the Royal Statistical Society. Series B **17**: 1-34.

Bramwell, D. I. and A. C. Hurlbert (1996). "Measurements of colour constancy using a forced choice matching technique." Perception **25**: 229-241.

Breakwell, G. M. (2006). Research methods in psychology. London, Sage.

Brenner, E. and F. W. Cornelissen (1998). "When is a background equivalent? Sparse chromatic context revisited." Vision Research **38**(12): 1789-1793.

Bridges, C. D. B. and S. Yoshikami (1969). "Distribution and evolution of visual pigments in Salmonid fishes." Vision Research **10**: 609-626.

Brindley, G. S. and D. I. Hamasaki (1966). "Histological Evidence Against the View that the Cat's Optic Nerve contains Centrifugal Fibres." Journal of Physiology **184**: 444-449.

Brodatz (1966). A texture library for artists and designers. New York, Dover Publications Inc.; .

Broerse, J., T. Vludsich, et al. (1999). "Colour at edges and colour spreading in McCullough effects." Vision Res. **39**: 1305-1320.

Brown, R. O. and D. I. A. MacLeod (1997). "Color appearance depends on the variance of surround colors." Current Biology **7**(11): 844-849.

Burt, A. M. (1993). Brain Stem and Cerebellum. Textbook of Neuroanatomy. Philadelphia, W. B. Saunders**:** 132-155.

Burt, A. M. (1993). Organization and development of the nervous system. Textbook of Neuroanatomy. Philadelphia, W. B. Saunders**:** 3-28.

Burt, A. M. (1993). Visual Pathways. Textbook of Neuroanatomy. Philadelphia, W. B. Saunders**:** 224-249.

Callaway, E. M. (2005). "Structure and function of parallel pathways in the primate early visual system." Journal of Physiology-London **566**(1): 13-19.

Carandini, M., J. B. Demb, et al. (2005). "Do we know what the early visual system does?" Journal of Neuroscience **25**(46): 10577-10597.

Carroll, J., C. Bialozynski, et al. (2000). "Estimates of L : M cone ratios from ERG flicker photometry and genetics." Investigative Ophthalmology & Visual Science **41**(4): S808-S808.

Carver, R. P. (1978). "The case against statistical significance testing." Harvard Educational Review **48**(3): 378-399.

Casagrande, V. A., G. Sáry, et al. (2005). "On the impact of attention and motor planning on the lateral geniculate nucleus." Progress in Brain Research **149**: 11-29.

Cavanagh, P. (1991). Vision at equiluminance. Vision and visual dysfunction: limits of vision. Cronly-Dillon, J. Kulikowski, Walsh and Murray, MacMillan**:** 234-250.

Chamberlin, T. C. (1965). "The method of Multiple Working Hypotheses." Science **148**(3671): 754-759.

Chaparro, A., I. Stromeyer, C.F., et al. (1993). "Colour is what the eye sees best." Nature **361**: 348-350.

Chevreul (1868). The laws of harmony and contrast of colours. London, Routledge.

Chichilnisky, E. J. and B. A. Wandell (1995). "Photoreceptor sensitivity changes explain color appearance shifts induced by large uniform backgrounds in dichoptic matching." Vision Research **35**(2): 239-254.

Chubb, C., G. Sperling, et al. (1989). "Texture Interactions Determine Perceived Contrast." Proceedings of the National Academy of Sciences of the United States of America **86**(23): 9631-9635.

Clarke, R. J. and P. D. R. Gamlin (1995). "The role of the pretectum in the pupillary light reflex." Basic and Clinical Perspectives in Vision Research: 149-159.

Clarke, S., V. Walsh, et al. (1998). "Colour constancy impairments in patients with lesions of the prestriate cortex." Experimental Brain Research **123**(1-2): 154-158.

Cohen, B. H. (2007). Explaining psychological statistics. Hoboken, N.J. Chichester, Wiley ; John Wiley distributor.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, N.J., L. Erlbaum Associates.

Cohen, J. (1994). "The Earth Is Round (P-Less-Than.05)." American Psychologist **49**(12): 997-1003.

Conway, B. R., S. Moeller, et al. (2007). "Specialized color modules in macaque extrastriate cortex." Neuron **56**(3): 560-573.

Cornelissen, F. W. and E. Brenner (1991). "Spatial Interactions in Color-Vision Depend on the Distance between Boundaries of Surfaces." Perception **20**(1): 99-99.

Cowey, A., P. Stoerig, et al. (1994). "Retinal Ganglion-Cells Labeled from the Pulvinar Nucleus in Macaque Monkeys." Neuroscience **61**(3): 691-705.

Cowey, A., P. Stoerig, et al. (1989). "Trans-Neuronal Retrograde Degeneration of Retinal Ganglion-Cells after Damage to Striate Cortex in Macaque Monkeys - Selective Loss of P-Beta Cells." Neuroscience **29**(1): 65-80.

Cribbie, R. A., J. A. Gruman, et al. (2004). "Recommendations for applying tests of equivalence." Journal of Clinical Psychology **60**(1): 1-10.

Cronbach, L. J. (1951). "Coefficient alpha and the internal structure of tests." Psychometrika **16**(3): 297-334.

Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleinig, A., Lo, J., and N. McMenamin, & Wilson, S "Statistical reform in psychology: Is anything changing?" Psychological Science **In press**.

Da Vinci, L. (1954). Theory of Colour. The Notebooks of Leonardo Da Vinci E. McCurdy, Plain Label books**:** 263-272.

Dacey, D. M. (1989). "Axon-Bearing Amacrine Cells of the Macaque Monkey Retina." Journal of Comparative Neurology **284**(2): 275-293.

Dacey, D. M. (2000). "Parallel pathways for spectral coding in primate retina." Annual Review of Neuroscience **23**: 743-775.

Dacey, D. M., H. W. Liao, et al. (2005). "Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN." Nature **433**(7027): 749-754.

Dacey, D. M. and O. S. Packer (2003). "Colour coding in the primate retina: diverse cell types and cone-specific circuitry." Current Opinion in Neurobiology **13**(4): 421-427.

Daniels, H. E. (1955). "Discussion on the paper by Dr. Box and Dr. Andersen." Journal of the Royal Statistical Society. Series B **17**: 27-29.

David, F. N. (1947). "A Power Function for Tests of Randomness in a Sequence of Alternatives." Biometrika **34**: 335-339.

Dayton, P. K. (1998). "Ecology - Reversal of the burden of proof in fisheries management." Science **279**(5352): 821-822.

De Valois, R. L., N. P. Cottaris, et al. (2000). "Some transformations of color information from lateral geniculate nucleus to striate cortex." PNAS **97**(9): 4997-5002.

DeAngelis, G. C., B. G. Cumming, et al. (1998). "Cortical area MT and the perception of stereoscopic depth." Nature **394**(6694): 677-680.

Derrington, A. M., J. Krauskopf, et al. (1984). "Chromatic mechanisms in lateral geniculate nucleus of Macaque." Journal of Physiology **357**: 241-265.

Desimone, R. and S. J. Schein (1987). "Visual Properties of Neurons in Area V4 of the Macaque - Sensitivity to Stimulus Form." Journal of Neurophysiology **57**(3): 835-868.

DeValois, R. (1981). "Citation Classic - Analysis of Response Patterns of LGN Cells." ENG TECH APPL SCI **1**: 16.

DeValois, R. L., J. Abramov, et al. (1966). "Analysis of response patterns of LGN cells." Journal of the Optical Society of America **56**: 966-977.

Dorr, S. and C. Neumeyer (2000). "Color constancy in goldfish: the limits." Journal of Comparative Physiology a-Sensory Neural and Behavioral Physiology **186**(9): 885-896.

Dulai, K. S., J. K. Bowmaker, et al. (1994). "Sequence divergence, polymorphism and evolution of the middle-wave and long-wave visual pigment genes of great apes and old-world monkeys." Vision Research **34**: 2483-2491.

Dumoulin, S. O., C. L. Baker, et al. (2003). "Cortical specialization for processing first- and second-order motion." Cerebral Cortex **13**(12): 1375-1385.

Eason, R. G., M. Oakley, et al. (1983). "Central Neural Influences on the Human Retina during Selective Attention." Physiological Psychology **11**(1): 18-28.

Editor. (2008). "Vision Research Guide for Authors " Retrieved 05/09/2008, 2008, from http://www.elsevier.com/wps/find/journaldescription.cws_home/263/authorinstructions.

Farnsworth, D. (1957). The Farnsworth-Munsell 100-Hue Test for the examination of color discrimination. Newburgh, NY, Macbeth division of Kollmorgen Instruments Corporation, Munsell Color.

Fechner, G. T. (1860). Elemente der Psychophysik. Leipzig, Breitkopf und Härtel.

Felleman, D. J. and D. C. Van Essen (1991). "Distributed Hierarchical Processing in the Primate Cerebral Cortex." Cerebral Cortex **1**(1): 1-47.

Felleman, D. J. and D. C. Vanessen (1987). "Receptive-Field Properties of Neurons in Area V3 of Macaque Monkey Extrastriate Cortex." Journal of Neurophysiology **57**(4): 889-920.

Felsten, G., L. A. Benevento, et al. (1983). "Opponent-Color Responses in Macaque Extrageniculate Visual Pathways - the Lateral Pulvinar." Brain Research **288**(1-2): 363-367.

Ferrera, V. P., T. A. Nealey, et al. (1994). "Responses in Macaque Visual Area V4 Following Inactivation of the Parvocellular and Magnocellular Lgn Pathways." Journal of Neuroscience **14**(4): 2080-2088.

Ferrera, V. P., K. K. Rudolph, et al. (1994). "Responses of Neurons in the Parietal and Temporal Visual Pathways during a Motion Task." Journal of Neuroscience **14**(10): 6171-6186.

Fisher, R. A. (1928). Statistical methods for research workers. Edinburgh ; London, Oliver and Boyd.

Fisher, R. A. (1935). The design of Experiments. London, Oliver & Boyd.

Fisher, R. A. (1955). "Statistical Methods and Scientific Induction." Journal of the Royal Statistical Society. Series B **17**: 69-78.

Fletcher, R. and J. Voke (1985). Defective Colour Vision: Fundamentals, Diagnosis and Management. Bristol, Adam Hilger.

Fodor, J. A. (1983). Modularity of Mind: An Essay on Faculty Psychology. Cambridge, MA, MIT Press.

Foster, D. H. and W. F. Bischof (1991). "Thresholds from Psychometric Functions - Superiority of Bootstrap to Incremental and Probit Variance Estimators." Psychological Bulletin **109**(1): 152-159.

Foster, D. H. and S. M. C. Nascimento (1994). "Relational color constancy from invariant cone-excitation ratios." Proceedings of the Royal Society of London - Series B **257**(1349): 115-121.

Foster, D. H., S. M. C. Nascimento, et al. (1997). Colour constancy from colour relations in the normal and colour-deficient observer. John Dalton's colour vision legacy. C. Dickinson, I. Murray and D. Garden. London, Taylor and Francis.

Gamlin, P. D. R., H. Y. Zhang, et al. (1995). "Luminance Neurons in the Pretectal Olivary Nucleus Mediate the Pupillary Light Reflex in the Rhesus-Monkey." Experimental Brain Research **106**(1): 177-180.

Gamlin, P. D. R., H. Y. Zhang, et al. (1998). "Pupil responses to stimulus color, structure and light flux increments in the rhesus monkey." Vision Research **38**(21): 3353-3358.

Gegenfurtner, K. R., D. C. Kiper, et al. (1997). "Functional properties of neurons in macaque area V3." Journal of Neurophysiology **77**(4): 1906-1923.

Gescheider, G. A. (1997). Psychophysics : the fundamentals. Mahwah, N.J., L. Erlbaum Associates.

Gibbs, T. and R. B. Lawson (1973). "Simultaneous brightness contrast in stereoscopic space." Vision Research **14**: 983-987.

Gigerenzer, G. (2004). "Mindless statistics." The Journal of Socio-Economics **33**: 587–606.

Gilchrist, A. (1977). "Perceived lightness depends on perceived spatial arrangement." Science **195**: 185-187.

Gnyubkin, V. F., S. L. Kondrashev, et al. (1975). "Constant Color Perception of Grey Toad." Biofizika **20**(4): 725-730.

Gobba, F. and A. Cavalleri (2003). "Color vision impairment in workers exposed to neurotoxic chemicals." Neurotoxicology **24**(4-5): 693-702.

Goethe (1840). Conditions of the Appearance of Colour. Goethe's Theory of Colours. London, John Murray**:** 81-106.

Gogel, W. C. and D. H. Mershon (1969). "Depth adjacency in simultaneous contrast." Perception & Psychophysics **5**(1): 13-17.

Goldstein, H. and M. J. R. Healy (1995). "The Graphical Presentation of a Collection of Means." Journal of the Royal Statistical Society A **158**(1): 175-177.

Goodale, M. A. and A. D. Milner (1992). "Separate Visual Pathways for Perception and Action." Trends in Neurosciences **15**(1): 20-25.

Goodale, M. A. and D. A. Westwood (2004). "An evolving view of duplex vision: separate but interacting cortical pathways for perception and action 2." Current Opinion in Neurobiology **14**(2): 203-211.

Green, D. M. and J. A. Swets (1988). Signal detection theory and psychophysics. Los Altos, Calif., Peninsular Publishing.

Grieve, K. L., C. Acuna, et al. (2000). "The primate pulvinar nuclei: vision and action." Trends in Neurosciences **23**(1): 35-39.

Gurnsey, R. and D. J. Fleet (2001). "Texture space." Vision Research **41**(6): 745-757.

H. Munsell, Boston 1905 (1905). A Color Notation. Boston, MA, Munsell color company.

Hadjikhani, N., A. K. Liu, et al. (1998). "Retinotopy and color sensitivity in human visual cortical area V8." Nature Neuroscience **1**(3): 235-241.

Haller, H. and S. Kraus (2002). "Misinterpretations of Significance: A Problem Students Share with Their Teachers?" Methods of Psychological Research **7**: 1-20.

Hammond, B. R., B. R. Wooten, et al. (1997). "Individual variations in the spatial profile of human macular pigment." Journal of the Optical Society of America a-Optics Image Science and Vision **14**(6): 1187-1196.

Hattar, S., H. W. Liao, et al. (2002). "Melanopsin-containing retinal. ganglion cells: Architecture, projections, and intrinsic photosensitivity." Science **295**(5557): 1065-1070.

Heywood, C. and A. Cowey (1998). "With color in mind." Nature Neuroscience **1**(3): 171-173.

Heywood, C. A., J. J. Nicholas, et al. (1998). "The effect of lesions to cortical areas V4 or AIT on pupillary responses to chromatic and achromatic stimuli in monkeys." Experimental Brain Research **122**(4): 475-480.

Hill, N. J. (1996-2009). Psignifit. Cambridge, Bootstrap software.

Hill, N. J. (2001). Testing Hypotheses about Psychometric Functions - an investigation of some confidence interval methods, their validity, and their use in the evaluation of optimal sampling strategies D. Phil., University of Oxford, UK. .

Hill, N. J. (2002-2008). "Psignifit FAQs." 2007, from http://www.bootstrap-software.org/psignifit/faq.php#toc-pfcmp.

Hinkle, D. A. and C. E. Connor (2001). "Disparity tuning in macaque area V4." Neuroreport **12**(2): 365-369.

Hinkle, D. A. and C. E. Connor (2005). "Quantitative characterization of disparity tuning in ventral pathway area V4." Journal of Neurophysiology **94**(4): 2726-2737.

Hochberg, J. E. and J. Beck (1954). "Apparent spatial arrangement and perceived brightness." Journal of Experimental Psychology **47**(4): 263-266.

Hoenig, J. M. and D. M. Heisey (2001). "The abuse of power: The pervasive fallacy of power calculations for data analysis." American Statistician **55**(1): 19-24.

Hopkins, K. D., B. R. Hopkins, et al. (1996). Basic statistics for the behavioral sciences. Boston, Allyn and Bacon.

Hunt, R. W. G. (2001). Measuring Colour. England, Fountain Pr. Ltd.

Hurlbert, A. C. (1998). Computational Models of Colour Constancy. Perceptual Constancy: Why things look as they do. V. Walsh and J. Kulikowski. Cambridge, Cambridge University Press**:** 283-322.

Hurlbert, A. C., D. I. Bramwell, et al. (1998). "Discrimination of cone contrast changes as evidence for colour constancy in cerebral achromatopsia." Experimental Brain Research **123**: 136-144.

Hurlbert, A. C. and T. Poggio (1989). A network for image segmentation using color. Neural Information Processing Systems. T. D. California, Morgan Kaufman Publishers**:** 297-303.

Hurlbert, A. C. and K. Wolf (2004). "Color contrast: a contributory mechanism to color constancy." Progress in Brain Research **144**: 147-160.

Hurlbert, A. C. H. (1998). "Illusions and reality-checking on the small screen." Perception **27**: 633.

Ingle, D. J. (1985). "The Goldfish as a Retinex Animal." Science **227**(4687): 651-654.

Intskirveli, Roinishvili, et al. (2002). "Experience-Dependent Color Constancy in Guppies (Poecilia reticulata)." Neural Plasticity **9**(3): 205-216.

Ioannidis, J. P. A. (2005). "Why most published research findings are false." Plos Medicine **2**(8): 696-701.

Ishihara, S. (1925). Tests for Colour-Blindness. Tokyo, Kanehara & Co.

Jacobs, G. H. (1981). Comparative color vision. New York, New York Academic Press.

Jacobs, G. H. (1993). "The distribution and nature of colouur vision among the mammals." Biol. Rev. **68**: 413-417.

Jacobs, G. H. and G. A. Williams (2001). "The prevalance of defective color vision in Old World monkeys and apes." Color Res. Appl. 26 **S123-S127**.

Jameson, D. and L. M. Hurvich (1960). "Opponent chromatic induction: experimental evaluation and theoretical account." Journal of the Optical Society of America A **51**: 46-53.

Jameson, D. and L. M. Hurvich (1960). "Opponnent Chromatic Induction: Experimental Evaluation and Theoretical Account." Journal of the Optical Society of America **51**(1): 46-53.

Jenness, J. W. and S. K. Shevell (1995). "Color Appearance with Sparse Chromatic Context." Vision Research **35**(6): 797-805.

Jewett, D. L. (2005). "What's wrong with single hypotheses." Scientist **19**(21): 10-10.

Joost, U., B. B. Lee, et al. (2002). "Lichtenberg's letter to Goethe on "Farbige Schatten" - Commentary." Color Research and Application **27**(4): 300-301.

Jordan, G. and J. D. Mollon (1993). "A Study of Women Heterozygous for Colour Deficiencies." Vision Research **33**(11): 1495-1508.

Joselevitch, C. and M. Kamermans (in press). "Retinal parallel pathways: Seeing with our inner fish." Vision Research.

Judd, D. B. (1943). "Colorblindness and the Detection of Camouflage." Science **97**(2529): 544-546.

Kamermans, M., D. A. Kraaij, et al. (1998). "The cone/horizontal cell network: A possible site for color constancy." Visual Neuroscience **15**(5): 787-797.

Kastner, S., K. A. Schneider, et al. (2006). Beyond a relay nucleus: neuroimaging views on the LGN. Progress in Brain Research. S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso and P. U. Tse, Elsevier. **155:** 125-144.

Kelly, D. H. (1983). "Spatiotemporal Variation of Chromatic and Achromatic Contrast Thresholds." Journal of the Optical Society of America **73**(6): 742-750.

Kennard, C., M. Lawden, et al. (1995). "Color Identification and Color Constancy Are Impaired in a Patient with Incomplete Achromatopsia Associated with Prestriate Cortical-Lesions." Proceedings of the Royal Society of London Series B-Biological Sciences **260**(1358): 169-175.

Kingsmith, P. E., S. S. Grigsby, et al. (1994). "Efficient and Unbiased Modifications of the Quest Threshold Method - Theory, Simulations, Experimental Evaluation and Practical Implementation." Vision Research **34**(7): 885-912.

Kirk, E. C. and S. E. L (2000). "Diets of fossil primates from the Fayum Depression of Egypt: a quantitative analysis of molar shearing." Journal of Human Evolution **40**: 203-229.

Kleiner, M. (2003). "MacOS-X Matlab timing bug: Interference with OS-X update - process?", from http://groups.yahoo.com/group/psychtoolbox/message/2227?threaded=1&var=1.

Knapp, T. R. (1998). "Comments on the Statistical Significance Testing Articles." Research in the Schools **5**: 39-41.

Kolb, H., K. A. Linberg, et al. (1992). "Neurons of the Human Retina - a Golgi-Study." Journal of Comparative Neurology **318**(2): 147-187.

Kraft, J. M. and D. H. Brainard (1997). "Mechanisms of color constancy under nearly natural viewing." Proceedings of the National Academy of Sciences USA **96**: 307-312.

Kraft, J. M. and D. H. P. o. t. N. A. o. S. U. Brainard, 96, 307-312. Download PDF. (1997). "Mechanisms of color constancy under nearly natural viewing." Proceedings of the National Academy of Sciences USA **96**: 307-312.

Krauskopf, J., Q. Zaidi, et al. (1986). "Mechanisms of simultaneous color induction." Journal of the Optical Society of America A **3**: 1752-1757.

L. Jiménez del Barco, J. A. D., J.R. Jiménez, M. Rubiño. (1995). "Considerations on the calibration of color displays assuming constant channel chromaticity." Color Res. Appl. **20**: 377-387.

Ladunga, K. (2001). "Relative luminosity generated by the colours of the CRT." Periodica polytechnica, mechanical engineering **45**(1): 59-64.

Lak, D. (2000, 6/9/2000). "Nepalese scorn Everest threat." 2008, from http://news.bbc.co.uk/1/hi/world/south_asia/912906.stm.

Langsrud, Y. (2003). "ANOVA for unbalanced data: Use Type II instead of Type III sums of squares." Statistics and Computing **13**(2): 163-167.

Lehmann, E. L. (1993). "The Fisher, Neyman-Pearson Theories of Testing Hypotheses - One Theory or 2." Journal of the American Statistical Association **88**(424): 1242-1249.

Lennie, P., J. Krauskopf, et al. (1990). "Chromatic Mechanisms in Striate Cortex of Macaque." Journal of Neuroscience **10**(2): 649-669.

Levene, H. (1960). Robust Tests for Equality of Variances. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. O. I. Palo Alto, Stanford University Press**:** 278-292.

Levin, J. (1998). "What If There Were No More Bickering About Statistical Significance Tests?" Research in the Schools **5**: 43-53.

Levine, M. W. and M. W. Levine (2000). Levine & Shefner's Fundamentals of sensation and perception. Oxford, Oxford University Press.

Levine, T. and J. Banas (2002). "One-tailed F-tests in communication research." Communication Monographs **69**(2): 132-143.

Lewsey, J. D., W. P. Gardiner, et al. (2001). "A study of type II and type III power for testing hypotheses from unbalanced factorial designs." Communications in Statistics-Simulation and Computation **30**(3): 597-609.

Lim, T. S. and W. Y. Loh (1996). "A comparison of tests of equality of variances." Computational Statistics & Data Analysis **22**(3): 287-301.

Lindman, H. R. (1992). Analysis of Variance in Experimental Design. New York, Springer-Verlag.

Lindsay, R. M. and A. S. C. Ehrenberg (1993). "The Design of Replicated Studies." American Statistician **47**(3): 217-228.

Loftus, G. R. (1991). "On the Tyranny of Hypothesis Testing in the Social Sciences." Contemporary Psychology **36**(2): 102-105.

Lucassen, M. P. and J. Walraven (1993). "Quantifying color constancy: evidence for nonlinear processing of cone-specific contrast." Vision Research **33**(5/6): 739-757.

Ludbrook, J. (2000). "Multiple Inferences using Confidence Intervals." Clinical and Experimental Pharmacology and Physiology **27**: 212-215.

Lyon, M. F. (1961). "Gene Action in the X-chromosome of the Mouse (Mus musculus L.)." Nature **190**: 372 - 373

MacLeod, D. I. A. and R. Boynton (1979). "Chromaticity diagram showing cone excitation by stimuli of equal luminance." Journal of the Optical Society of America **69**(8): 1183-1187.

Malinas, G. and J. Bigelow (2004). Simpson's Paradox. The Stanford Encyclopedia of Philosophy. E. N. Zalta. Stanford, The Metaphysics Research Lab Center for the Study of Language and information Stanford University Stanford, CA 94305-4115.

Mangun, G. R., J. C. Hansen, et al. (1986). "Electroretinograms Reveal No Evidence for Centrifugal Modulation of Retinal Inputs during Selective Attention in Man." Psychophysiology **23**(2): 156-165.

Marks, L. E. (1974). Sensory processes: the new psychophysics. New York,, Academic Press.

Marrocco, R. T. and R. H. Li (1977). "Monkey Superior Colliculus - Properties of Single Cells and Their Afferent Inputs." Journal of Neurophysiology **40**(4): 844-860.

Masland, R. H. (2001). "The fundamental plan of the retina." Nature Neuroscience **4**(9): 877-886.

Mathworks. (2006). "Matlab Programming - techniques for improving performance." from http://www.mathworks.com/access/helpdesk/help/techdoc/matlab.html.

Maunsell, J. H. R., T. A. Nealey, et al. (1990). "Magnocellular and Parvocellular Contributions to Responses in the Middle Temporal Visual Area (Mt) of the Macaque Monkey." Journal of Neuroscience **10**(10): 3323-3334.

Maunsell, J. H. R. and D. C. Vanessen (1983). "Functional-Properties of Neurons in Middle Temporal Visual Area of the Macaque Monkey .2. Binocular Interactions and Sensitivity to Binocular Disparity." Journal of Neurophysiology **49**(5): 1148-1167.

McAlonan, K., J. Cavanaugh, et al. (2008). "Guarding the gateway to cortex with attention in visual thalamus." Nature **456**: 391-394

McCann, J. J. (2003). Calculating appearances in complex and simple images. J. D. Mollon, J. Pokorny and K. Knoblauch. Oxford, Oxford university press**:** 231-238.

McDonald, C. G. and C. W. Hawryshyn (1999). "Latencies and discharge patterns of color-opponent neurons in the rainbow trout optic tectum." Vision Research **39**(17): 2795-2799.

McQueen, R. A. and C. Knussen (2006). Pearson Education.

Meehl, P. E. (1967). "Theory testing in psychology and physics: a methodological paradox." Philosophy of Science **34**: 103-115.

Meehl, P. E. (1997). The Problem Is Epistemology, Not Statistics: Replace Significance Tests by Confidence Intervals and Quantify Accuracy of Risky Numerical Predictions. What If There Were No Significance Tests? . S. A. M. L. L. Harlow, & J. H. Steiger New Jersey, Lawrence Erlbaum**:** 393–425.

Mershon, D. H. (1971). "Relative contributions of depth and directional adjacency to simultaneous whiteness contrast." Vision Research **12**: 969-979.

Milton, K. (1993). "Diet and Primate Evolution." Scientific American **269**: 86-93.

Minolta, K. Chroma Meter CS-100 Instruction Manual, Konica Minolta.

Moller, P. and A. C. Hurlbert (1996). "Psychophysical evidence for fast region-based segmentation processes in motion and color." Proceedings of the National Academy of Sciences of the United States of America **93**(14): 7421-7426.

Mollon (1989). "Tho' she kneel'd in that place where they grew... The uses and origins of primate colour vision." Journal of Experimental Biology **146**: 21-38.

Mollon, J. (2006). "Monge - The Verriest Lecture, Lyon, July 2005." Visual Neuroscience **23**(3-4): 297-309.

Mollon, J. D. (1999). Specifying, generating and measuring colours. Vision Research A Practical Guide to Laboratory Methods. R. H. S. Carpenter and J. G. Robson. Oxford, UK, Oxford University Press**:** 106-127.

Mollon, J. D., J. K. Bowmaker, et al. (1984). "Variations of Color-Vision in a New World Primate Can Be Explained by Polymorphism of Retinal Photopigments." Proceedings of the Royal Society of London Series B-Biological Sciences **222**(1228): 373-399.

Morgan, M. J., A. Adam, et al. (1992). "Dichromates detect color-camouflaged objects that are not detected by trichromates." Proc. Roy. Soc. Lond. Series B-Biol. Sci. **248**: 291-295.

Morgan, M. J., A. Adam, et al. (1992). "Dicromats detect color camouflaged objects that are not detected by trichromats." Proc. R. Soc. Lond. B **248**: 291-295.

Motulsky, H. and A. Christopolous (2003). Fitting Models to Biological Data using Linear and Nonlinear Regression : A practical guide to Curve Fitting. La Jolla, GraphPad Software.

Muliak, S. A., N. S. Raju, et al. (1994). There is a time and place for significance testing. What If There Were No Significance Tests? L. Lavoie Harlow, S. A. Muliak and J. A. Steiger, Lawrence Erlbaum Associates**:** 65-117.

Mullen, K. T. and G. T. Plant (1986). "Color and Luminance Vision in Human Optic Neuritis." Brain **109**: 1-13.

Muller, T., W. Kuhn, et al. (1999). "Colour vision abnormalities and movement time in Parkinson's disease." European Journal of Neurology **6**(6): 711-715.

Muma, J. R. (1993). "The Need for Replication." Journal of Speech and Hearing Research **36**: 927-930.

Nascimento, S. M. C., F. P. Ferreira, et al. (2002). "Statistics of spatial cone-excitation ratios in natural scenes." Journal of the Optical Society of America a-Optics Image Science and Vision **19**(8): 1484-1490.

Nascimento, S. M. C. and D. H. Foster (2001). "Detecting changes of spatial cone-excitation ratios in dichoptic viewing." Vision Research **41**(20): 2601-2606.

Nathans, J. (1999). "The evolution and physiology of human color vision: Insights from molecular genetic studies of visual pigments." Neuron **24**(2): 299-312.

Nathans, J., T. P. Piantanida, et al. (1986). "Molecular-Genetics of Inherited Variation in Human Color-Vision." Science **232**(4747): 203-210.

Nayatani, Y., K. Takahama, et al. (1988). "Physiological causes of individual variations in color-matching functions." Color Research and Application **13**(5): 289 - 297.

Neitz, M. and J. Neitz (2000). "Molecular genetics of color vision and color vision defects." Archives of Ophthalmology **118**(5): 691-700.

Neumeyer, C. (1998). Comparative aspects of colour constancy. Perceptual constancy: why things look as they do. V. Walsh and J. Kulikowski. Cambridge, UK, Cambridge University Press.

Neumeyer, C., S. Dorr, et al. (2002). "Colour constancy in goldfish and man: influence of surround size and lightness." Perception **31**(2): 171-187.

Northcutt, R. G. and J. H. Kaas (1995). "The Emergence and Evolution of Mammalian Neocortex." Trends in Neurosciences **18**(9): 373-379.

Nvidia (2002). Digital vibrance control. Santa Clara, CA, NVIDIA Corporation 2701 San Tomas expressway Santa Clara, CA 95050.

O'Connor, D. H., M. M. Fukui, et al. (2002). "Attention modulates responses in the human lateral geniculate nucleus." Nature Neuroscience **5**(11): 1203-1209.

O'Keefe, L. P. and J. A. Movshon (1998). "Processing of first- and second-order motion signals by neurons in area MT of the macaque monkey." Visual Neuroscience **15**(2): 305-317.

Ohzawa, I. (1998). "Mechanisms of stereoscopic vision: the disparity energy model." Current Opinion in Neurobiology **8**(4): 509-515.

Olveczky, B. P., S. A. Baccus, et al. (2003). "Segregation of object and background motion in the retina." Nature **423**(6938): 401-408.

Onishi, A., S. Koike, et al. (1999). "Vision - dichromatism in macaque monkeys." Nature **402**: 139-140.

Ord, T., R. Hillerbrand, et al. (2009). Probing the Improbable: Methodological Challenges for Risks with Low Probabilities and High Stakes. ArXiv.

Parker, A. J. (2007). "Binocular depth perception and the cerebral cortex." Nature Reviews Neuroscience **8**(5): 379-391.

Parraga, C. A., T. Troscianko, et al. (2000). "The human visual system is optimised for processing the spatial information in natural visual images." Current Biology **10**(1): 35-38.

Peat, J. and B. Barton (2005). Medical Statistics, Blackwell publishing.

Pelli, D. G. and B. Farell (1999). Psychophysical methods, or how to measure a threshold and why. Vision Research: A Practical Guide to Laboratory Methods. R. H. S. Carpenter and J. G. Robson. New York, Oxford University Press**:** 129-136.

Peterson, B. B. and D. M. Dacey (1998). "Morphology of human retinal ganglion cells with intraretinal axon collaterals." Visual Neuroscience **15**(2): 377-387.

Photo-Research. (2009). "PR-650 SpectraScan® Colorimeter." from http://www.photoresearch.com/current/pr650.asp.

Pitz, G. F. (1978). "Hypothesis Testing and Comparison of Imprecise Hypotheses." Psychological Bulletin **85**(4): 794-809.

Platt (1964). "Strong Inference." Science **146**(1946): 347-353.

Pokorny, J. and V. C. Smith (1985). "L/M cone ratios and the null point of the perceptual red/green opponent system." Die Farbe **34**: 53-57.

Pokorny, J., V. C. Smith, et al. (1987). "Aging of the Human Lens." Applied Optics **26**(8): 1437-1440.

Popper, K. (2007). The logic of Scientific Discovery. New York, Routledge Classics.

Quirk, J. A., D. R. Fish, et al. (1995). "First Seizures Associated with Playing Electronic Screen Games - a Community-Based Study in Great-Britain." Annals of Neurology **37**(6): 733-737.

Rao, A. R. and G. L. Lohse (1993). "Towards a Texture Naming System - Identifying Relevant Dimensions of Texture." Visualization 93, Proceedings: 220 & 423.

Read, J. (2005). "Early computational processing in binocular vision and depth perception." Progress in Biophysics and Molecular Biology **87**: 77-108.

Read, J. C. A., A. J. Parker, et al. (2002). "A simple model accounts for the response of disparity-tuned V1 neurons to anticorrelated images." Visual Neuroscience **19**(6): 735-753.

Rees, G., G. Kreiman, et al. (2002). "Neural correlates of consciousness in humans." Nature Reviews Neuroscience **3**(4): 261-270.

Regan, D. and K. I. Beverley (1984). "Figure-Ground Segregation by Motion Contrast and by Luminance Contrast." Journal of the Optical Society of America a-Optics Image Science and Vision **1**(5): 433-442.

Reid, R. C. and R. M. Shapley (1992). "Spatial Structure of Cone Inputs to Receptive-Fields in Primate Lateral Geniculate-Nucleus." Nature **356**(6371): 716-718.

Réperant, J. and A. Gallego (1976). "Fibres centrifuges dans la rétine humaine." Arch. Anat. Microscop. Morphol. Exp. **65**: 103-120.

Réperant, J., R. Ward, et al. (2006). "The centrifugal visual system of vertebrates: A comparative analysis of its functional anatomical organization." Brain Research Reviews **52**(1): 1-57.

Rhodes, R. (1986). The making of the atomic bomb. New York, Simon & Schuster paperbacks.

Riecansky, I., A. Thiele, et al. (2005). "Chromatic sensitivity of neurones in area MT of the anaesthetised macaque monkey compared to human motion perception." Experimental Brain Research **167**(4): 504-525.

Rinner, O. and K. R. Gegenfurtner (2000). "Time course of chromatic adaptation for color appearance and discrimination." Vision Research **40**: 1813-1826.

Robson, C. (1994). Experiment, design and statistics in psychology. Harmondsworth, England, Penguin.

Robson, T. (1999). Topics in computerized visual-stimulus generation. Vision Research. A practical guide to laboratory methods. R. H. S. Carpenter and J. G. Robson. Oxford, Oxford University Press.

Rosenthal, R. (1979). "The File Drawer Problem and Tolerance for Null Results." Psychological Bulletin **86**(3): 638-641.

Rubin, J. M. and W. A. Richards (1982). "Color-Vision and Image Intensities - When Are Changes Material." Biological Cybernetics **45**(3): 215-226.

Rudd, M. E. and K. F. Arrington (2001). "Darkness filling-in: a neural model of darkness induction." Vision Research **41**: 3649-3662.

Schein, S. J. and R. Desimone (1990). "Spectral Properties of V4 Neurons in the Macaque." Journal of Neuroscience **10**(10): 3369-3389.

Schiller, P. H., J. G. Malpeli, et al. (1979). "Composition of Geniculostriate Input to Superior Colliculus of the Rhesus-Monkey." Journal of Neurophysiology **42**(4): 1124-1133.

Schirillo, J. A. and S. K. Shevell (2000). "Role of perceptual organization in chromatic induction." Journal of the Optical Society of America a-Optics Image Science and Vision **17**(2): 244-254.

Schneider, K. A., M. C. Richter, et al. (2004). "Retinotopic organization and functional subdivisions of the human lateral geniculate nucleus: A high-resolution functional magnetic resonance imaging study." Journal of Neuroscience **24**(41): 8975-8985.

Schuirmann, D. J. (1987). "A Comparison of the 2 One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability." Journal of Pharmacokinetics and Biopharmaceutics **15**(6): 657-680.

Seime, L. and J. Y. Hardeberg (2002). Characterisation of LCD and DLP projection displays. IS&T/SID's Color Imaging Conference, Scottsdale, Arizona.

Shady, S. and D. I. A. MacLeod (2002). "Color from invisible patterns." Nature Neuroscience **5**(8): 729-730.

Shaw, R. G. and T. Mitchell-Olds (1993). "Anova for unbalanced data: an overview." Ecology **74**(6): 1638-1645.

Shepherd, A. J. (1997). "A vector model of colour contrast in a cone-excitation colour space." Perception **26**(4): 455-470.

Sherman, S. M. and C. Koch (1986). "The Control of Retinogeniculate Transmission in the Mammalian Lateral Geniculate-Nucleus." Experimental Brain Research **63**(1): 1-20.

Shevell, S. K. and P. R. Miller (1996). "Color perception with test and adapting lights perceived in different depth planes." Vision research **36**(7): 949-954.

Shevell, S. K. and J. P. Wei (2000). "A central mechanism of chromatic contrast." Vision Research **40**(23): 3173-3180.

Shipp, S. (2003). "The functional logic of cortico-pulvinar connections." Philosophical Transactions of the Royal Society B-Biological Sciences **358**(1438): 1605-1624.

Simon, A., E. Martin-Martinelli, et al. (2001). "Confirmation of the retinopetal/centrifugal nature of the tyrosine hydroxylase-immunoreactive fibers of the retina and optic nerve in the weaver mouse." Developmental Brain Research **127**(1): 87-93.

Simons, D. J. and C. F. Chabris (1999). "Gorillas in our midst: sustained inattentional blindness for dynamic events." Perception **28**(9): 1059-1074.

Skrabanek, P. and J. McCormick (1998). Follies and fallacies in medicine. Whithorn, England, Tarragon Press.

Smeets, J. B. J. and E. Brenner (1994). "The Difference between the Perception of Absolute and Relative Motion - a Reaction-Time Study." Vision Research **34**(2): 191-195.

Smith, A. T., P. L. Cotton, et al. (2009). "Dissociating Vision and Visual Attention in the Human Pulvinar." Journal of Neurophysiology **101**(2): 917-925.

Smith, R. W. (1997). Visual Hypothesis testing with Confidence Intervals. SAS Users Group International 22, San Diego, California.

Smith, V. C. and J. Pokorny (1975). Vision Res. **15**: 161-172.

Smithson, H. E. (2005). "Sensory, computational and cognitive components of human colour constancy." Philosophical Transactions of the Royal Society B-Biological Sciences **360**(1458): 1329-1346.

Solomon, S. G. and P. Lennie (2007). "The machinery of colour vision." Nature Reviews Neuroscience **8**(4): 276-286.

Stabell, B. and U. Stabell (1998). "Chromatic rod-cone inter action during dark adaptation." Journal of the Optical Society of America a-Optics Image Science and Vision **15**(11): 2809-2815.

Stewart, E. C. (1959). "The Gelb Effect." Journal of Experimental Psychology **57**(4): 235-242.

Stoerig, P. and A. Cowey (1992). "Wavelength Discrimination in Blindsight." Brain **115**: 425-444.

Stoughton, C. M. and B. R. Conway (2008). "Neural basis for unique hues." Current Biology **18**(16): 698-699.

Student (1908). "The probable distribution of a mean." Biometrika **6**: 1-25.

Sumner, P. and J. D. Mollon (2000). "Catarrhine photopigments are optimized for detecting targets against a foliage background." Journal of experimental biology **203**: 1963-1986.

Sumner, P. and J. D. Mollon (2000). "Chromaticity as a signal of ripeness in fruits taken by primates." Journal of experimental biology **203**: 1987-2000.

Takkouche, B., C. Cadarso-Suarez, et al. (1999). "Evaluation of Old and New Tests of Heterogeneity in Epidemiologic Meta-Analysis." American Journal of Epidemiology **150**(2): 206-215.

Taylor, M. M., S. M. Forbes, et al. (1983). "Pest Reduces Bias in Forced Choice Psychophysics." Journal of the Acoustical Society of America **74**(5): 1367-1374.

Thompson (1998). "Statistical Significance and Effect Size Reporting: Portrait of a Possible Future." Research in the Schools **5**(2): 33-38.

Thompson, B. (2001, 2/26/01). "402 Citations Questioning the Indiscriminate Use of Null Hypothesis Significance Tests in Observational Studies ", 2008.

Thouless (1931). "Phenomenal regression to the “real” object (I)." British Journal of Psychology **21**: 331-359.

Tobimatsu, S., Y. M. Zhang, et al. (1999). "Chromatic sensitive epilepsy: A variant of photosensitive epilepsy." Annals of Neurology **45**(6): 790-793.

Tolias, A. S., G. A. Keliris, et al. (2005). "Neurons in macaque area V4 acquire directional tuning after adaptation to motion stimuli." Nature Neuroscience **8**(5): 591-593.

Tootell, R. B. H., K. Nelissen, et al. (2004). "Search for color 'center(s)' in macaque visual cortex." Cerebral Cortex **14**(4): 353-363.

Trevor-Roper, P. (1988). The World Through Blunted Sight. London, Penguin Books, Ltd.

Trujillo-Ortiz, A. and R. Hernandez-Walls (2003). Levenetest: Levene's test for homogeneity of variances.

Umeda, K., S. Tanabe, et al. (2007). "Representation of stereoscopic depth based on relative disparity in macaque area V4." Journal of Neurophysiology **98**(1): 241-252.

Usai, C., G. M. Ratto, et al. (1991). "2 Systems of Branching Axons in Monkeys Retina." Journal of Comparative Neurology **308**(2): 149-161.

Vaina, L. M. and A. Cowey (1996). "Impairment of the perception of second order motion but not first order motion in a patient with unilateral focal brain damage." Proceedings of the Royal Society B-Biological Sciences **263**(1374): 1225-1232.

Vaina, L. M., N. Makris, et al. (1998). "The selective impairment of the perception of first-order motion by unilateral cortical brain damage." Visual Neuroscience **15**(2): 333-348.

Valberg, A. (1974). "Color Induction - Dependence on Luminance, Purity, and Dominant or Complementary Wavelength of Inducing Stimuli." Journal of the Optical Society of America **64**(11): 1531-1540.

Vanduffel, W., R. B. H. Tootell, et al. (2000). "Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system." Cerebral Cortex **10**(2): 109-126.

Velleman, P. F. and L. Wilkinson (1993). "Nominal, Ordinal, Interval, and Ratio Typologies Are Misleading." American Statistician **47**(1): 65-72.

von Kries, J. (1902). Chromatic Adaptation, Festschrift der Albrecht-Ludwigs-Universitat, pp. 145-158 Sources of Color Science, 1970. MacAdam. Cambridge, MIT Press.

Vos (1978). "Colorimetric and photometric properties of a 2° fundamental observer " Color Research and Application **3**(3): 125-128.

Wachtler, T., T. D. Albright, et al. (2001). "Nonlocal interactions in color perception: Nonlinear processing of chromatic signals from remote inducers." Vision Research **41**(12): 1535-1546.

Wachtler, T., T. J. Sejnowski, et al. (2003). "Representation of color stimuli in awake macaque primary visual cortex." Neuron **37**(4): 681-691.

Wallach, H. (1948). "Brightness constancy and the nature of achromatic colors." Journal of Experimental Psychology **38**: 310-324.

Walraven, J. (1973). "Spatial characteristics of chromatic induction; the segregation of lateral effects from straylight artefacts." Vision Research **13**: 1739-1753.

Ware, C. and W. B. Cowan (1982). "Changes in Perceived Color Due to Chromatic Interactions." Vision Research **22**(11): 1353-1362.

Watanabe, M., H. Tanaka, et al. (2002). "Disparity-selective neurons in area V4 of macaque monkeys." Journal of Neurophysiology **87**(4): 1960-1973.

Watson, A. B. (1983). "QUEST: a Bayesian adaptive psychometric method." Perception & Psychophysics **33**(2): 113-120.

Webster, M. A. (1996). "Human color vision and its adaptation." Network: Computation in Neural Systems **7**: 587-634.

Webster, M. A. and J. D. Mollon (1991). "Changes in colour appearance following post-receptoral adaptation." Nature **349**: 235-238.

Webster, M. A. and J. D. Mollon (1994). "The Influence of Contrast Adaptation on Color Appearance." Vision Research **34**(15): 1993-2020.

Weller, R. E. and J. H. Kaas (1989). "Parameters Affecting the Loss of Ganglion-Cells of the Retina Following Ablations of Striate Cortex in Primates." Visual Neuroscience **3**(4): 327-349.

Werner, A., L. T. Sharpe, et al. (2000). "Asymmetries in the time-course of chromatic adaptation and the significance of contrast." Vision Research **40**: 1101-1113.

White, C. T. and C. L. White (1995). "Reflections on visual evoked cortical potentials and selective attention: methodological and historical." International Journal of Neuroscience **80**: 13-30.

Whittle, P. and P. D. C. Challands (1969). "The effect of background luminance on the brightness of flashes." Vision Research **9**: 1095-1110.

Wichmann, F. A. and N. J. Hill (2001). "The psychometric function: II. Bootstrap-based confidence intervals and sampling." Perception & Psychophysics **63**(8): 1314-1329.

Wilhelm, B. J., H. Wilhelm, et al. (2002). "Pupil response components: studies in patients with Parinaud's syndrome." Brain **125**: 2296-2307.

Williams, C. H. (1915). "Nagel's anomaloscope for testing color-vision." Transactions of the American Ophthalmological Society **14**(1): 161-165.

Wolf, K. (2002). "Visual ecology: Coloured fruit is what the eye sees best." Current Biology **12**(7): R253-R255.

Wolf, K. and A. C. Hurlbert (2003). "Color contrast: a contributory mechanism to color constancy." Progress in Brain Research **144**(147-160).

Wolf, K. and A. C. H. Hurlbert (2002). "The contribution of local and global cone-contrasts to colour appearance: a Retinex-like model." Human vision and Electronic Imaging VII **4662**: 286-297.

Wolter, J. R. and R. R. Knoblich (1965). "Pathway of centrifugal fibres in the human optic nerve, chiasm and tract." British Journal of Opthalmology **49**: 246-250.

Wunderlich, K., K. A. Schneider, et al. (2005). "Neural correlates of binocular rivalry in the human lateral geniculate nucleus." Nature Neuroscience **8**(11): 1595-1602.

Wyble, D. R. and H. Zhang (2003). Colorimetric characterization model for DLP projectors. IS&T/SID's Color Imaging Conference, Scottsdale, Arizona.

Wyszecki, G. and W. S. Styles (1982). Color science: concepts and methods, quantitative data and formula. New York, Wiley.

Yamauchi, Y., D. R. Williams, et al. (2002). "What determines unique yellow, L/M cone ratio or visual experience?" Aic: 9th Congress of the International Colour Association **4421**: 275-278.

Yang, J. N. and L. T. Maloney (2001). "Illuminant cues in surface color perception: tests of three candidate cues." Vision Research **41**(20): 2581-2600.

Yuille, A. and D. Kersten (2006). "Vision as Bayesian inference: analysis by synthesis?" Trends in Cognitive Sciences **10**(7): 301-308.

Zachar, G., A. Schrott, et al. (2008). "Context-dependent prey avoidance in chicks persists following complete telencephalectomy." Brain Research Bulletin **76**(3): 289-292.

Zaidi, Q. (2000). Color and brightness induction: from Mach bands to three-dimensional configurations. Color Vision - From Genes to Perception. K. R. Gegenfurtner and L. T. Sharpe. Cambridge, UK, Cambridge University Press**:** 317-343.

Zaidi, Q., B. Yoshimi, et al. (1992). "Lateral Interactions within Color Mechanisms in Simultaneous Induced Contrast." Vision Research **32**(9): 1695-1707.

Zaidi, Q., B. Yoshimi, et al. (1991). "Influence of Shape and Perimeter Length on Induced Color Contrast." Journal of the Optical Society of America a-Optics Image Science and Vision **8**(11): 1810-1817.

Zaidi, Q. and N. Zipser (1993). "Induced Contrast from Radial Patterns." Vision Research **33**(9): 1281-1286.

Zeki, S. (1980). "The Representation of Colors in the Cerebral-Cortex." Nature **284**(5755): 412-418.

Zeki, S. (1990). "A Century of Cerebral-Achromatopsia." Brain **113**: 1721-1777.

Zeki, S. (1991). "Cerebral Akinetopsia (Visual-Motion Blindness) - a Review." Brain **114**: 811-824.

Zeki, S. M. (1973). "Color Coding in Rhesus-Monkey Prestriate Cortex." Brain Research **53**(2): 422-427.

Zipser, K., V. A. F. Lamme, et al. (1996). "Contextual modulation in primary visual cortex." Journal of Neuroscience **16**(22): 7376-7389.