A Calculation of Colours

towards the automatic creation
of
graphical user interface colour schemes

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Philosophy
in
Computer Science
at
Massey University, Palmerston North,
New Zealand.

Giovanni S. Moretti

2010

Dedicated to Edris and Serafino, my parents.

Abstract

Interface colour scheme design is complex, but important. Most software allows users to choose the colours of single items individually and out of context, but does not acknowledge colour schemes or aid in their design. Creating colour schemes by picking individual colours can be time-consuming, error-prone, and frustrating, and the results are often mediocre, especially for those without colour design skills. Further, as colour harmony arises from the interactions between all of the coloured elements, anticipating the overall effect of changing the colour of any single element can be difficult.

This research explores the feasibility of extending artistic colour harmony models to include factors pertinent to user interface design. An extended colour harmony model is proposed and used as the basis for an objective function that can algorithmically assess the colour relationships in an interface colour scheme. Its assessments have been found to agree well with human evaluations and have been used as part of a process to automatically create harmonious and usable interface colour schemes.

A three stage process for the design of interface colour schemes is described. In the first stage, the designer specifies, in broad terms and without requiring colour design expertise, colouring constraints such as grouping and distinguishability that are needed to ensure that the colouring of interface elements reflects their semantics.

The second stage is an optimisation process that chooses colour relationships to satisfy the competing requirements of harmonious colour usage, any designer-specified constraints, and readability. It produces sets of coordinates that constitute abstract colour schemes: they define only relationships between coloured items, not real colours.

In the third and final stage, a user interactively maps an abstract scheme to one or more real colour schemes. The colours can be fine-tuned as a set (but not altered individually), to allow for such "soft" factors as personal, contextual and cultural considerations, while preserving the integrity of the design embodied in the abstract scheme. The colours in the displayed interface are updated continuously, so users can interactively explore a large number of colour schemes, all of which have readable text, distinguishable controls, and conform to the principles of colour harmony.

Experimental trials using a proof-of-concept implementation called the Colour Harmoniser have been used to evaluate a method of holistic colour adjustment and the resulting colour schemes. The results indicate that the holistic controls are easy to understand and effective, and that the automatically produced colour schemes, prior to fine-tuning, are comparable in quality to many manually created schemes, and after fine-tuning, are generally better.

By designing schemes that incorporate colouring constraints specified by the user prior to scheme creation, and enabling the user to interactively fine-tune the schemes after creation, there is no need to specify or incorporate the subtle and not well understood factors that determine whether any particular set of colours is "suitable". Instead, the approach used produces broadly harmonious schemes, and defers to the developer in the choice of the final colours.

Acknowledgements

I would like to thank those who have contributed to this research project:

my family, for their support and understanding during what has turned out to be a longer than expected “absence”;

those who helped with the experimental work, especially: Dr Heather Meikle and the students from Palmerston North Girls High School; the staff at TeManawa Museum and Science Centre; staff and students at Massey University and at the UCOL School of Photography, Art & Design;

Dr Alasdair Noble, Dr Geoff Jones and Dr Jonathan Godfrey for their assistance with the statistics;

Dr Chris Phillips and to Dr Greg Allan their insightful questions and suggestions, and to my colleagues in the School of Engineering and Advanced Technology for their encouragement; and

my supervisors, Mr Paul Lyons and Dr Stephen Marsland, for their suggestions, and their patience and support with the so many aspects that, looking back, are now so obvious.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

We live surrounded by colour, both natural and artificial. While the colour combinations in the natural world are generally harmonious (i.e. pleasing to the eye), the colour schemes of manufactured goods, and more recently, virtual objects such as interfaces, are not necessarily as pleasant. Advances in technology have made it possible for designers to use any colour they choose, but the wide range of possible colourings has not simplified the selection of harmonious colour schemes.

Some individuals seem to be easily able to pick sets of colours that, when seen together, provide a pleasing effect, whether for clothes, interior design or graphic design. These individuals, often said to be "artistic", seem to have an innate understanding of how to choose sets of colours that look good together. There is another, larger, group who find the task of selecting colours – especially groups of colours – daunting, and who approach the colour selection process more or less randomly. For those without intuitive ability or training, selecting groups of harmonious colours can be a frustrating experience as, when their schemes are less appealing than they would like, they are at a loss to know what to do about it.

It is tempting to refer to terms like *colour harmony* or *harmonious colour schemes* and assume that their meaning is evident, and that there will be agreement about whether a particular colour scheme is harmonious (or not). This would be a mistake for two reasons. Firstly, as will be discussed in the next chapter, colour is a perceptual phenomenon and is therefore subjective. Secondly, *harmony* relates to the pleasing coherence of an arrangement of parts. The harmony of an arrangement, including its colour harmony, is related to aesthetics, which is also subjective and therefore not amenable to a simple objective definition. Until a more thorough discussion in section 2.9, the working definition of colour harmony will be that of Burchett (1991): "a set of colours that are pleasing when seen together".

Experience suggests that the colour scheme of a graphical user interface is rarely considered during the design of a software artifact. Usually, it is not regarded as particularly important, and the default colours of the interface components are used. Even when creating a website (for which there are no default colours), those with little aptitude for colour design rarely seek expert guidance. The results of such a cavalier attitude to colour scheme design are seldom appealing and can be ill-suited to the intended task. As will be shown, a colour scheme can have subtle and important effects on the viewers.

While those with limited colour design ability may recognise their own shortcomings, three factors have impeded the transfer of knowledge from those who can intuitively create pleasing schemes to those who cannot: the perceived unimportance of colour scheme design; the lack of ready access to appropriate colour design expertise, and the difficulty that those with intuitive ability can have when trying to elucidate how or why they choose the colours they do – they simply "know". Nevertheless, over the years, heuristics for the selection of harmonious colours scheme have come out of the artistic community. The principles can be learnt, but doing so takes time and practice, preferably with the guidance of a mentor. However, due to limits on time and the limited availability of readily accessible colour design expertise, software developers currently have little or no understanding of the principles of colour design.

This thesis is an investigation into whether the guidelines for colour scheme design from the artistic community can be applied, and if necessary extended, to allow the creation of a software tool to automate the design of pleasing user interface colour schemes. The intended users of such a tool are software and web developers who need to create interfaces for web sites and desktop software without the assistance of a graphic designer. The intent is definitely not to replace artists and designers – such individuals can have great perception and skill and should always be consulted if possible. However, there are a very large number of individuals creating software and web sites for whom professional assistance is not readily available. The aim is to help these developers produce better colour schemes than they would without assistance.

1.1 Colouring user interfaces

With the growth of the World Wide Web and the widespread availability of integrated software development environments and web site design tools, graphical user interfaces are now routinely designed by individuals, either at home or within small businesses. Working independently or in small businesses with limited staffing, these developers frequently have no colour design expertise, either personally or available in-house.

When it comes to selecting a colour scheme, they most often use the default colours or attempt to create a colour scheme themselves. Possible reasons for this include: the importance of the colour scheme is underestimated; contracting out the graphic design is time-consuming and expensive; and lastly, the independent and creative nature of software developers means that they can be loath to admit that they cannot create a scheme themselves: "you just open the colour selector and select a new colour for each item; how hard can it be?"

While developers may be competent at writing software or providing the functionality behind websites, their technical skills may not include the ability to colour their interfaces well. This can lead to frustration as, even after many iterations, their attempts can still be unsatisfactory. Such developers can be technically sophisticated but artistically naïve. They are capable of recognising a colour scheme that appeals, but are not necessarily capable of creating such a scheme, as MacIntyre (1991) succinctly states: *"typical users make far better critics than designers"*.

A personal motivation

To provide the perspective of an experienced software developer who was naïve about colour, I will relate my own experience when first designing a web site.

Creating the structure for the site itself was straightforward, but choosing a colour scheme was much more difficult than anticipated. Being reluctant to simply copy the scheme of an existing site, I tried to create a colour scheme myself. The schemes were garish and amateurish, even to one with little design expertise. When continued modifications did not seem to help, I turned to books on art and graphic design. The chapters on colour harmony described various methods of selecting complementary and split-complementary colours using a colour wheel. Puzzlingly, when the colours chosen using these methods were applied to the site, the result was just as garish and looked just as bad (although in a different way) as my initial uninformed attempts.

Evidently, either the guidelines were incorrect (which seemed improbable as they had been written by successful artists who presumably knew what they were talking about), or I was applying them incorrectly. While this seemed more likely, it was difficult to understand how such simple guidelines – along the lines of "use complementary colours, those opposite on a colour wheel" – could be misinterpreted.

Having followed the guidelines, I was perplexed. Eventually, I found my error, which was related to my not understanding the balancing of colour strength against area, concepts that will be introduced in the next chapter. I did not understand – and the books did not tell me – that colour wheels are typically printed with only the fullest intensity (or saturation) colours that the printing process is capable of producing, and harmonious colour schemes generally involve subtler, less intense versions of these colours, particularly when the areas to be coloured are large. Attempting to apply this insight has resulted in this thesis.

My experience encapsulates the problem. Colour design appears deceptively easy. For those without intuitive ability, it isn't. Without appropriate guidance, it is possible to waste a large amount of time and still only achieve mediocre results. One of the aims of this thesis is a system that could help improve the quality of the resulting colour schemes without requiring a significant time investment.

1.1.1 Colour selection for computer interfaces is important

Before considering why developers may find colour scheme creation difficult, it is necessary to show that the colour scheme of a user interface is sufficiently important to justify spending time improving its appearance.

Colour is highly valued as a means of personal expression. The huge range of colour schemes for clothing, interior design and web interfaces testify to this. We react subconsciously to colour (Kobayashi, 1981). It can have both physiological effects and psychological associations. Specific colours can evoke particular feelings. These associations are particularly important commercially. Changing the colour of an object's packaging, or more recently, a web site colouring, can have immediate impact on sales (van Geel, 2006), and the aesthetics can affect the perceived trustworthiness of e-commerce sites (Wang and Emurian, 2005; Karvonen, 2000).

The colouring of an interface also affects its usability. Inappropriate colour choices can make the interface less comfortable to use, especially for prolonged periods, and directly affect the readability of text and visual search times (Hilscher, 2005). For some industries, such as air traffic control and security screening, these factors can have safety implications.

While individual colour preferences are subjective, overall there are trends in the use and associations of colour that appear to transcend this. If this were not the case, it would be impossible to create colour schemes with a mass-market appeal, such as the widely-used white and bright leafy green colour scheme associated with freshness and cleanliness.

Interface colour selection is important. Whether used as a means of personal expression via web logs and personal web sites, commercially for business e-commerce sites, or for computer applications, it affects our experience while using a web site or software package, and our perceptions of its creator.

1.1.2 Colour selection is hard

The number of possible colourings of an interface is extremely large. For c distinguishable colours and n interface elements there are c^n possible colour schemes, not all distinct. To put this into context, an interface with seven colourable elements and using the now-standard 24 bit colour selector enables a user to choose $16,777,216^7$ different colour schemes. While the number of possible colour schemes is huge, the number of "good" colour schemes is very much smaller. This has been validated experimentally in this thesis, with the results presented in chapter 5. In an experimental trial where human assessors ranked colour schemes created by randomly choosing interface element colours, on a scale of "terrible" to "excellent", the average was "terrible". It would appear that in order to select good colour schemes, a method is necessary.

User interface colour scheme selection really is a difficult problem, both from an absolute perspective (the space is very large with very few pleasing combinations) and from the perspective of a user (the number of possible colour combinations is overwhelming, and the tools usually available for colour selection make choosing colours an error-prone and protracted process).

1.1.3 Colour selection tools are naïve

Most of the colour selectors available in current software are simplistic. They enable the user to select colours individually with no cognizance of their use in the context of a colour scheme. Typically, the colours will be chosen using a modal colour selector in which the final colour being chosen is shown on a small swatch, out of context, and usually selected using a much smaller area than that in which the colour will be seen. The appearance of a region of colour is affected by its size and the colours of nearby objects and the background (Fairchild, 1998). Therefore, a colour chosen out of context (such as the swatch in the colour selection dialog) will almost certainly look different when applied to an object in the interface. This necessitates invoking the colour selection dialog again, and again, and again.

Advances in desktop computing power and display technology have improved the quality of displayed images, but have not had any impact on the methods used to select colour schemes. The problem of colour selection and its current "piecemeal" approach will not be significantly altered by the advances in technology: faster CPUs and better graphics accelerators will not simplify the selection of good colour schemes using the current one-colour-at-a-time approach.

In applications that are oriented towards visual design, predefined colour schemes as well as individual colour selection may be available. This is a significant improvement over individual colour selection, but the overall approach is limited, with the user trying to find a colour scheme from a small and often fixed sampling of all possible colour schemes. It may also be possible to change the dominant colours used in the scheme. This is preferable to the "out-of-context" selection of individual colours, but there are severe limitations imposed by the inability to significantly vary the nature of the "canned" schemes. A method that treats the colour selection process holistically rather than piecemeal or "selection from the catalogue" is necessary.

It is also worth noting that colour schemes for interfaces may require the consideration of factors not necessary in other fields of colour design, such as the incorporation of differing colour preferences related to globalisation, and pragmatic considerations the designer may wish to include. Any method used to create interface colour schemes must be sufficiently flexible to allow for the incorporation of such "soft" factors.

1.2 Guidelines for the design of interface colour schemes

The overall appeal of a set of colours is not the sum of the appeal of the individual colours. It is the holistic impression of the overall scene, and is affected by visual interactions between the colours of objects (Birren, 1961), in which their size, spatial relationships and the user's personal colour preference all play a part.

There are various heuristics from the world of art that provide guidelines for selecting harmonious sets of colours. Most use simple geometric relationships on a colour wheel to suggest palettes that would generally be considered harmonious. However, the colour wheel is a one dimensional[1] slice of colour space, which is three dimensional. Therefore only colours from a restricted region of the colour space can be incorporated into the colour schemes produced using such methods. That was the reason the colour schemes I chose for my website were so garish. They were the only colours displayed on the colour wheel: the more subtle, lower-saturation colours were simply omitted.

[1] the basic colour wheel shows only one colour dimension: the pure colours (the hues). The colouring is as though a strip which varies in colour lengthwise but not across, has been stretched and wrapped into a circle, so although the wheel itself is two-dimensional, it only shows one dimension of colour. Hue variations, such as tints and less intense colours are not shown. The dimensions of colour space are discussed at the start of the next chapter.

There are general guidelines in the HCI (Human-Computer Interaction) literature on both the design of interfaces and, to a much lesser extent, on how to colour an interface. Unfortunately, the use of colour in interface design is rarely given more than a few pages. The five pages devoted to colour (out of 579 content pages) in Shneiderman et al. (2009) is not atypical. Given that the colouring affects every interface component, and is seen by every user of an interface, the scant treatments are inadequate to give sufficient guidance to interface designers.

1.2.1 User interfaces are not art

As the design of colour schemes is not covered in depth in the HCI literature, the world of art seems an obvious place to look. However, interfaces are not art. They differ in three distinct ways. Firstly and most importantly, interfaces are primarily intended to convey information, not to create an aesthetic effect. The colour of objects within any interface (especially informational text) cannot be coloured solely to satisfy artistic goals. The readability of text is important, and sometimes, as in medical instrumentation or air traffic control displays, critically so. Secondly, the appeal of a colour scheme is related to both the layout of objects and the colours used for those objects. The placement of elements within an interface may well be predetermined, and even if it is not, cannot be arbitrarily changed to satisfy artistic goals. Lastly, there are semantic implications relating to colour choice. Users may well draw inferences about the relatedness of items in an interface based on the similarity in their colouring. Therefore, while it is desirable for an interface colour scheme to be visually appealing, usability must not be subjugated to aesthetics.

1.2.2 Constraints on the use of colour in interface colour schemes

It is unlikely that every element in an interface will have a unique colour. This would completely ignore the semantics of the interface: that some interface elements are related, either by proximity, or by function. An interface where every control (e.g. captions, buttons, items of text) is a different colour is possible, but the disjointed colouring would be distracting, and distractions impair usability.

In interface design, colour is used in several distinct ways: to generate a visually pleasing appearance; to appeal to a specific target audience or to evoke a particular emotional or associative response; and to reflect the underlying functionality of the interface by enabling groups of controls to be visually grouped or partitioned by their colour.

The first two factors (visual appeal and evoked emotions and associations) are interrelated and very complex to define – whether or not a particular colour scheme is appropriate (even if theoretically harmonious) depends on the intended audience. The acceptability of a colour scheme relates to style and fashion, both of which are virtually impossible to define and are constantly changing.

Any attempt to codify these factors would require a system that was self-updating to avoid becoming obsolete. One possibility would be to query online resources and use image analysis to determine the colour schemes currently being used in a particular domain (e.g. fashion, cosmetics, interior design or automotive design). The extracted colour schemes could be used to guide the selection of new schemes conforming to the current fashion. While this could form an interesting project, it is outside the scope of the research described here.

1.2.3 The design process: client and designer

To understand the stages in the design of a colour scheme, it is appropriate to review the process of a client working with a colour design consultant. In an initial meeting, the client would typically describe the desired result, either by colour or intended effect (e.g. restful, contemplative, energising). The consultant would then prepare a selection of designs for the client's consideration, with these designs being used as the basis for further discussion. After several iterations, possibly requiring completely new designs, a result satisfying the client's criteria, both expressed and tacit, can usually be found.

This process incorporates both the domain expertise of the designer and the subjective preferences of the client. This last point – satisfying the client – is obviously critical, but also highly problematic. It is quite common for clients not to have a clear idea of what they want, and to rely on the designer to present them with a range of alternatives. This dichotomy between wanting a design, and not knowing what that design looks like, is not unusual. When faced with an overwhelming number of choices (a common situation where style or design are concerned), humans seem to be much more able to indicate when an option is *not* what they want, than they are of being able to specify what they *do* want. Overall, people are happier when the number of options is limited (Schwartz, 2005).

The iterative design process outlined above could be mirrored in a software tool. "Clients" (software or website developers) could provide the information needed during the iterative development of a pleasing colour scheme, even if they (initially) had little idea of what such a colour scheme might look like. As transparency in how the schemes are created is not necessary, a subsystem that produces schemes – a colour design "oracle" – could stand in for the human design consultant.

In order to replicate this process in software, the tool would need to contain or have access to:

1. the interface that needs to be coloured, and a means of updating the colours displayed.

2. a mechanism for the developer to express their initial preferences (even if ill-defined) for the appearance of the interface colour scheme.

3. a means of creating suitable "quality" colour schemes.
   This is necessary as the developer may have little or no colour design expertise, and no clear idea of what an acceptable colour scheme will look like.

4. a means for the developer to view the created schemes in a realistic setting, ideally in-situ, on the actual interface.

5. a means for the developer to either suggest or make changes.

An interface created by an IDE[2] or displayable as a website satisfies the first criterion (having an object to colour, and one that can have its colour updated) and the fourth, that of being able to see any colour scheme in-situ, by updating the interface colour scheme.

This leaves items two (a way for the developer to specify their preferences), three (creating good interface colour schemes) and five (a means to suggest or make changes). The first and the last (specifying preferences and incorporating suggestions or changes) would need to be expressed in non-specialist language as the target audience (developers) are unlikely to be familiar with artistic concepts. The only remaining factor is the automatic creation of interface colour schemes. This is the most challenging, but there are artistic heuristics for creating harmonious colour schemes and these could provide a starting point.

All the required elements necessary to mirror the human-to-human design consultancy process seem amenable to algorithmic decomposition.

1.3 The thesis

It seems possible that the lack of colour design knowledge shown by many application and website developers could be compensated for by a tool that incorporated a formal model of colour harmony, which it would use to facilitate the automatic selection of harmonious sets of colours. This research is intended to establish whether or not it is feasible to produce such a tool and to determine whether or not the colour schemes that could be produced by such a tool meet some minimum standard of quality. The tool produced in the research would be designed for proof of concept, not for production use.

To the extent that an interface is made up of a number of individual components, an automatic colouring system is obliged to select colours individually. To the extent that colour harmony is a holistic percept, the individual colours are less important than the relationships between them. Therefore, an automatic colouring system that is intended to produce harmonious colour schemes should treat the colours of interface items as an integrated whole. On the face of it, these requirements are contradictory, but it is possible to devise an approach to interface colouring that conforms to both of them by treating the creation of a harmonious colour scheme as a two-phase process. The first phase defines the relationships between the coloured items but not their actual colours. The second phase allows the assignment of colours while preserving the relationships defined in phase one. The result is a colour scheme whose colours satisfy a model of colour harmony, but whose actual colouring is the result of human adjustment.

> *The system would initially define the relationships between colours in an abstract colour space, and use these to create a family of raw, or unadjusted, colour schemes, each of which could subsequently be mapped to many sets of*

[2] IDE (Integrated Development Environment) such as found in modern software development tools, e.g. Microsoft's *Visual Studio*™, CodeGear's *Delphi*™ (previously Borland's Delphi) or Adobe's *Dreamweaver*™ for web design.

real colours. This mapping could be accomplished by incorporating a simple mechanism to allow users to holistically tweak the colouring to their preference. This research would be considered successful if users considered the quality of the raw schemes to be comparable to the quality of schemes created by human developers unskilled in colour design, and if users considered the final (tweaked) schemes to be of significantly higher quality than the raw schemes.

To validate this thesis, it is necessary to

- extend the artistic models of colour harmony so that they are:
  - applicable to user interface design,
  - susceptible to being described and used algorithmically,
  - incorporate user preferences with regard to GUI characteristics and overall colouring.
- develop a software tool that incorporates the extended colour harmony model, and can be used to create user interface colour schemes and facilitate their adjustment.
- experimentally validate the extended colour harmony model through user evaluations of the resulting colour schemes with those created using a more conventional method of colour selection.
- test whether users find the holistic method of colour selection and adjustment easier to use than conventional methods.

There are a number of features that could be included in a tool for automating the production of pleasing interface and website colour schemes for graphic user interfaces and web sites, but have been excluded: textures and gradients, maintaining colour harmony over multiple pages, and the manual modification of the colours of individual interface elements. The system envisaged will be oriented to create colour schemes for use in typical viewing conditions and suited to the colour preferences of adults, not children, whose preferences differ (see 2.1.4, p16). None of these factors is essential to the evaluation of the thesis. Providing assistance to colourblind users is not a focus of the research.

The next chapter covers the appropriate background (i.e. human colour perception, the use and effects of colour, theories of colour harmony, and systems for interface colour scheme design). The remaining chapters discuss a conceptualisation of an automatic interface colouring system, a software prototype to show the feasibility of the conceptualisation, and the results of experimental evaluation the prototype and its resulting colour schemes.

Chapter 2

Background and related work

This thesis aims to test the idea that the automatic design of harmonious colour schemes is possible. Such a statement raises some questions. What is colour? What is meant by a harmonious colour scheme? Why might it be desirable to have a harmonious colour scheme on a user interface? Is it reasonable to think that it might be possible to create such schemes automatically?

Defining colour harmony precisely is problematic, as will become evident in section 2.9, but a *harmonious colour scheme* can be understood as one that is aesthetically pleasing to the viewer. No absolute measure of colour harmony has been discovered – it appears to be largely subjective – but there are commonly accepted guidelines that allow the derivation of sets of colours that, when applied appropriately, many users find pleasing.

Before discussing the guidelines themselves, it is necessary to cover three other topics, in order to enable the design principles to be understood. The first is the impact of colour choice on a viewer, the second is a survey of how colour can be used in interface design, and the third is how physiological considerations have affected computing technology and the choice of colour spaces used to select harmonious sets of colours. At the conclusion of that discussion, sufficient background will have been covered to detail methods of actually selecting the sets of colours themselves, and how others have applied these techniques in user interface design.

There are several terms that are essential when discussing colour. To specify a colour using a name is far too ambiguous: which blue *exactly* is "sky blue"? There are many blues, some light, some dark, some greyish-blue. In order to clearly define a colour, it is sufficient to specify three characteristics (Birren, 1969b; Fairchild, 1998):

hue: a measure of the dominant colour of a particular object. It is difficult to define hue without using relative definitions, but it can be understood by thinking of hue as that characteristic of a colour that gives rise to its name (e.g. red, green, orange, purple).

lightness: the perceived amount of light from a colour object or area compared to black and white. For example, a full strength (high saturation) yellow appears to be almost as light as white, and therefore has a high lightness. A pure blue appears darker and therefore has a lower lightness. Generally, lightness increases in the order black, blue, yellow, white.The term *value* is an older term for lightness and is still widely-used.

saturation: this is defined in Fairchild (1998) as *"colorfulness of an area judged in proportion to its brightness"*. However, it may be easier to think of saturation in terms of the strength or purity of a colour. The saturation of a colour can be reduced without altering the hue by adding white to give *tints*, or adding black to give *shades*, or adding grey, the resultant colours sometimes being called *tones*. The older term *chroma* is widely used as an alternative to saturation. While Fairchild notes a technical distinction between the two, in common usage they are used interchangeably (and will be in this thesis).

The difference between lightness and brightness: some authors confuse these. Lightness is relative, whereas brightness is absolute. Lightness is a relative measure of the amount of light from the brightest colour in a scene to the darkest, whereas brightness is an absolute measure of the amount of light being emitted (or reflected) by an object or area.

Consider the text on this page[1]. When viewed inside on a sunny day, the text appears black, the page white. If the same page is taken outside, the page will be very much brighter, but the page is still seen as white and the text as black. The lightness has not changed, but the brightness will be orders of magnitude higher.

When referring to colours on a display, it is almost always lightness that is meant. Areas where the backlight is completely blocked on an LCD display are perceived as black, but it is incorrect to describe them as having a brightness of zero because, although they emit no light, they do reflect some ambient light; this "flare" sets the zero on the lightness scale. Pixels where red, green and blue are completely unblocked appear white. All measures of lightness on a display are relative to these two limits.

2.1 The impact of colour and colour schemes

Colour surrounds us and is immensely important on both personal and commercial levels. Even among those who profess not to care about colour, few would be happy to wear randomly coloured clothes, especially when applying for a job (this is fortunate, as the perceived competence of male job applicants can be significantly affected by the colour of their clothing (Damhorst and Reed, 1986)).

2.1.1 Colour: associations and emotions

The connection between colours and emotions is an important factor during colour selection. The literature on the emotional aspects of colour is extensive, and although not always in agreement, it is clear that any method of colour scheme design must allow

[1] this example is based on that of Fairchild (1998)

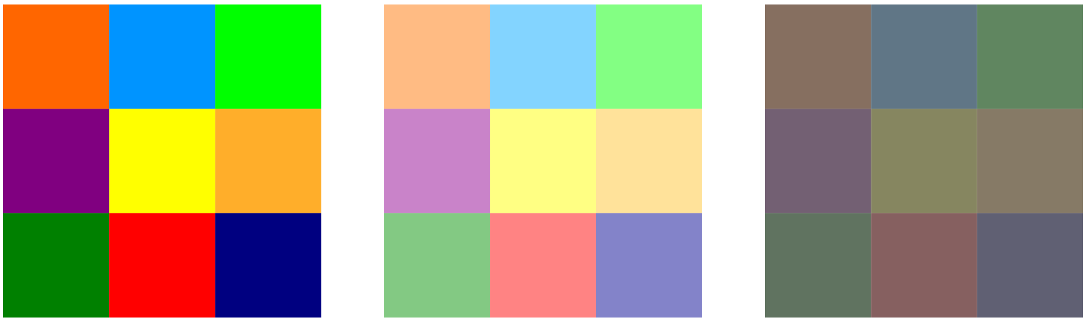


Figure 2.1: Nine colours, all the same hue, but with differing lightness and saturations in each image. Left: high saturation; centre, lighter (tinted – mixed with white); right, desaturated and darker (shades). The emotional reaction to each of the three figures is likely to be quite different. All three redrawn from Hornung (2005).

the emotive effects to be taken into account. This section explores the emotive effects of single colours, while the next section considers the effects of colours in combination.

Valdez and Mehrabian (1995) present a comprehensive overview of the connection between colour and emotions. They review both the physiological and psychological aspects and then present experimental results based on a pleasure–arousal–dominance (PAD) emotion model. In terms of colour preference: *"blue, blue-green, green, purple-blue, red-purple, and purple were the most pleasant; whereas yellow, green-yellow, and red-yellow were the least pleasant; with red being rated at an intermediate value of pleasantness"*. Generally, light and less saturated colours were more pleasant and darker colours more dominant. Note that these are reactions to individual colours, not to colour schemes. Importantly, Valdez and Mehrabian suggest that effects of confounding hue with brightness and saturation could explain inconsistencies in earlier studies. They state:

> *"The present studies provided highly consistent evidence of color brightness and saturation to emotional reactions. In comparison, relationships of hue to emotions were surprisingly weak."*

Both sexes reacted in similar ways to lightness and saturation differences, the reaction to differing hues was also independent of gender, and: *"consistent support was obtained for proposed hypotheses relating pleasure to hue (or wavelength)"*.

An illustration of the effects of varying hue and lightness can be seen in figure 2.1, derived from Hornung (2005). The hues are identical in all three images, but the emotional reaction to each is likely to be quite different.

The effect of colours on college students has been surveyed by Kaya and Epps (2004) and, in contrast to the results of Valdez and Mehrabian (1995), yellow was not disliked:

> *"The color yellow was seen to be lively, energetic, and brought feelings of happiness, and excitement because it was associated with the sun, blooming flowers, and summer time.*
>
> ...

> *The achromatic colors evoked the most feelings of depression, sadness, boredom, confusion, tiredness, loneliness, anger, fear, and death." – Kaya and Epps (2004).*

Consistent reactions from younger children were reported by Boyatzis and Varghese (1994) and, like Kaya and Epps, they noted that emotional responses are strongly linked to past experience.

Light of certain wavelengths does have physiological effects, as is indicated by the widely-used blue-light phototherapy method of treating neonatal infants with jaundice (Kaiser, 1984; Ennever, 1990). Amato and Inaebnit (1991) found significant differences between the effects of blue and green light, with the use of blue light requiring shorter treatment times than green. Other colours also have effects. Red is commonly thought to be a stimulating colour. This has been borne out experimentally. In inkblot tests, the effect of red was clearly different from that of other colours: *"According to the results, normals, patients with neurotic disorders, and borderline patients expressed love and anger significantly more often in response to cards containing red colors." – Leichsenring (2004).*

Pink, although the same hue as red, evokes quite a different reaction – it has a significant calming effect. Listed in the US National Criminal Justice Reference Services web site[2] (ref: NCJ069079) is a paper by Schauss (1979), where the use of pink walls in law enforcement holding cells at the U.S. Naval Correctional Center, Seattle, and also at the Santa Clara County Jail, was found to significantly reduce aggression. Valdez and Mehrabian (1995), in referring to this work, noted that all three characteristics of the pink used ("Baker–Miller" pink[3]): light, low saturation and of a purple-red hue, were found by Valdez and Mehrabian to reduce arousal, consistent with the effect found by Schauss. The interior colour of a building also affects the cooperativity of children in schools (Read et al., 1999).

Some hues appear to be special, with two colours, red and blue, singled out for particular study: red (Leichsenring, 2004; Sarapik, 1997; Caivano and López, 2003); and blue, widely held to be the most common favourite colour in Western cultures and studied as a topic unto itself (Mendoza, 2004; Paramei, 2005). It has not always been so favoured: Pastoureau (2001) points out that blue was reviled in ancient times, and it has only risen to popularity relatively recently, over the last five centuries.

2.1.2 Emotive impact of multiple colours – colour schemes

Colours are rarely used in isolation. More commonly, several are used in combination and the emotional reaction to the overall combination is not the aggregated or average appeal of the individual colours. When coloured objects are close together, the overall effect may be harmonious or not. The reasons for this appear to be subtle and complex and will be discussed in section 2.9, but the desire for a simple way of selecting pleasing colour combinations and achieving a desired emotional reaction has led to many books containing colour palettes.

[2] `http://www.ncjrs.gov/App/Publications/abstract.aspx?ID=69079` - accessed May 21, 2010.

[3] An sRGB approximation to Baker-Miller pink is (255,145,175)

A typical example is Sawahata (1999) in which colour schemes are grouped by theme or emotion, and include notes on the application of the palette to each of graphic design, interior design and fine art. Each "themed" palette (e.g. "elegant") is illustrated with using differing types of colour scheme: monochromatic, primary colours, complementary, split-complementary, analogous, clashing colours, and neutral (hue with grey). Other publications with colour scheme ideas for use in colour design with categorisations (e.g. emotional intent) include Krause (2002), Barker (1999), Hornung (2005), Wright (1995), Foster (2004) and Lleonart (2009).

The consideration of aesthetics extends into the engineering evaluation in the field of Kansei engineering: the intentional engineering of emotional aspects into product design. Wu and Chuang (2000) report on a linguistically-based experiment to determine guidelines for the design of personal web pages intended to evoke a particular feeling in the viewer. Kansei engineering is also discussed in Grimsaeth (2007), Chuang and Ma (2001) and Yamakazi and Kondo (1999).

2.1.3 Colour and culture

The use of colour needs to be appropriate for the context in which it is used, whether it is used to convey information, for decoration, or to evoke an impression or feeling. Whatever the context – personal appearance, car colour, interior decorating, an advertisement, or the interface of a computing system – it is preferable that any colour-related emotional connotations or associations be intentional, not accidental.

The context may vary, but many colour associations and the feelings evoked by particular colours or sets of colours appear to be independent of culture. Kobayashi (1987) derived the "Colour Image Scale" – a colour space derived from linguistic analysis of emotive terms, The research was conducted in Japan and yet, examining the words associated with particular colour sets, one could imagine that the survey had been conducted in New Zealand. In a comparison of emotive terms to colours, very significant agreement was found between the Tzeltal-speaking inhabitants of Chipas (the southern-most part of Mexico) and Northern American participants (D'Andrade and Egan, 1974). The Tzeltal language is one of the Mayan family of languages with the authors stating: *"the results of the color-emotion test indicate that within the limits of translation equivalence color chips and emotion terms show very similar patterns of association in both cultures"*. Similarly, Zollinger (1988) confirms that naming of the basic terms of Hering (1964)[4] – red, green, blue and yellow – appears to be stable against cross-cultural influence, a result supported by the correlations found between English and Chinese speakers in the placement of colours in ten colour-emotion scales (Ou et al., 2004a,b), and a more extensive study involving participants from six countries (Ou et al., 2008).

[4] Hering's basic colours and their derivation will be discussed in section 2.3.1.

2.1.4 Colour scheme appeal is age-specific

The appeal of colour schemes is not only culture-dependent, but varies with age (Nemcsics, 2009). In a survey of 3000 participants, Palffy (1976) found that the appeal for loud and discordant colours peaked at age 12–14, and was uncommon in adults. Both Morriss et al. (1982) and Palffy (1976) have found in experimental trials that vivid discordant schemes are preferred by children.

The visual appeal of bright strong colours has been validated experimentally (Camgöez et al., 2002) with undergraduates as subjects. A preference for strongly saturated foreground colours when tested against a fully saturated background was found, with a blue background being much preferred. A statistically significant preference for strongly coloured and high brightness[5] samples was found, with no support for classical colour harmony hue arrangements. This is not surprising, as the use of complementary or analogous colour schemes would not normally be used against fully saturated backgrounds. The preference for a blue background may be purely because of a liking for blue, or possibly because a blue background was the least bright and therefore provided a less intense impression, or both. Fully saturated backgrounds are rarely pleasant and, as noted by Sagawa (1999), for natural scenes, overall saturation is inversely related to comfort.

The vivid schemes preferred by children indicate that differing heuristics would be required to create harmonious colour schemes for adults and children. The system being envisaged in this thesis is oriented towards adults, and while acknowledging children's preferences, they will not be considered further.

2.2 The use of colour in graphical user interfaces

Colour in a computer interface can be used to convey information or, as humans can have quite strong emotional reactions to colour, to evoke an emotional response. If an interface colour scheme is "appropriate", it may go unnoticed, but if not, it is desirable that any emotional reaction be positive, not negative. It is therefore important to consider how an interface colour scheme may affect both the user and the functioning of an interface.

Tools for the creation of more visually-oriented media, such as web pages or presentations (e.g. Microsoft's *PowerPoint*™), may incorporate a set of predefined schemes. However, the variety of schemes provided can be rather constrained. The other common approach, changing the colours of individual elements to create a colour scheme, allows completely unconstrained creation, but is time-consuming and the results, if the user is not artistically-oriented, can be mediocre.

Software design environments, such as CodeGear's *Delphi*™ or Microsoft's *Visual Studio*™ will normally use the system default colours for all the interface elements, and while it is possible to change each element, an integrated approach to colour scheme management is unlikely to be available. The developer may change the colour of any

[5] Brightness is the appropriate term here, as the experimenters used the *Hue-Saturation-Brightness* colour space to choose the test colours.

individual interface element (or not) as they see fit, without assistance from the design environment. As most developers are not also colour designers, few choose to alter the defaults, with the result that almost all software products use the default interface colouring, which, while being inoffensive, also tends to be rather bland.

Preferences for (and against) individual colours and complete schemes apply to interfaces as well as real-world objects. Dissatisfaction with the widespread uniformity of software applications has led to the development of *skinning* – the ability to apply new themes to an existing interface to enable the colours of the interface elements and fonts to be easily modified without recompilation or access to the original source code. The wide popularity of this approach can be seen in the development of web sites and companies devoted to creating and distributing themes for different products. For example, the Wordpress blogging system (`http://www.wordpress.org`) has a page devoted to themes (`http://www.wordpress.org/extend/themes`) that lists 927 themes with 6,879,400 downloads[6], and twenty-eight sites with themes for sale `http://www.wordpress.org/extend/themes/commercial`, one example being `http://www.ithemes.com`. The Firefox web browser is also capable of being “skinned”. Some themes are listed on the site `https://addons.mozilla.org/en-US/firefox/browse/type:2`, with the most popular theme on this page having 47,920 downloads per week. As is evident, personalising interface colour schemes is popular.

2.2.1 Aesthetics affect usability

User interface displays can be made to conform to usability style guidelines, such as those given by Gulliksen and Sandblad (1995), and while these may ensure accessible functionality, they do not address the issue of personal aesthetics. There are generic guidelines for producing schemes suitable for particular target user groups, but ideally, the interface colour scheme should be configurable by the end-user.

The importance of aesthetics in user interface design has been demonstrated experimentally by Tractinsky et al. (2000). Users who were presented with variations of the same interface with identical functionality, but differing aesthetics, found the interfaces whose aesthetics they preferred to be significantly more usable than those with lower aesthetic appeal. This corroborated the earlier results of Kurosu and Kashimura (1995). Aesthetically pleasing products are not only perceived as easier to use, but are also used more frequently (Jordan, 1998). Aesthetics is therefore one of the criteria used to determine product appeal and longevity (Bloch, 1995).

2.2.2 Colour affects usability

The many and varied effects of the colours in an interface complicate the selection of colours for use in user interface colour schemes. One of the few detailed discussions about the use of colour in an interface from a perspective other than aesthetics is that of Travis (1991, pp112-144) who details four ways that colour can be used when designing a user interface. These four points can be summarised as:

[6] The URLs accessed and counts on August 12, 2009.

to provide a realistically rendered colour display: the need for realism of the rendered colours would usually apply only to some regions within an interface, for example those used to render photographs or graphic designs intended for print. The accurate control of colour within these areas would necessitate the use of a colour management system with appropriate profiles, and depending on the precision needed, appropriate control of the viewing conditions. This research is not intended to develop automatic colouring systems for interfaces that incorporate photo-realistic images.

to provide formatting: colour can be used to lay out the display and partition the interface elements, or to draw attention to particular elements.

to encode information: colour can be used to encode meaning, but if care is not taken, the colours chosen can seem arbitrary or can clash with the viewer's preconceptions. For example, because of prior conditioning, a bright red area as a system status indicator would probably be interpreted as indicating that something was wrong – red is not an ideal colour to indicate an "all ok" condition. Equally, a green background would be a poor choice for an alarm condition. Because prior colour associations (red implies stop/alarm/danger, green implies go/all-ok) give an implicit meaning to coloured areas, it would be poor design to attempt to override these expectations.

Visualisation systems generalise the mappings of colours to meanings, where colour is used to indicate the value of some attribute. In such cases, there are usually insufficient predefined associations, and a legend is necessary to indicate which colour is associated with a value or range of values. This mapping of colours to meanings is problematic: *"because they do have a natural visual hierarchy, varying shades of gray show varying quantities better than color."* whereas *"color often generates graphical puzzles. Despite our experiences with the spectrum in science textbooks and rainbows, the mind's eye does not readily give a visual ordering to colours" – Tufte (2001, p154).*

to add aesthetics: although the primary use of a graphical user interface is to convey information, this does not preclude it from also being aesthetically pleasing. The use of appropriate colours can affect both the visual appeal of an interface and the subconscious evoked emotional response. The importance given to the aesthetics of an interface varies. It could be thought, given that aesthetics affects perceived usability, that interface colouring would be regarded as important, but not all agree. Nielsen and Hahir suggests that website designers' should not waste time allowing colour customisation in the design of a web site. Referring to the colour scheme:

> *"Better to offer one design that is as readable as possible for the majority of users and let users who need to modify colors do so through their browser controls." – Nielsen and Hahir (2002, p231).*

The suggested colour customisation is technically problematic. Overriding the web site designer's options involves dynamically rewriting either the HTML or style sheets, both of which are well beyond what could be expected of a typical user. It is both simpler for the user, and more likely to result in a coherent appearance, if colour scheme options are "designed-in" by the website designer.

The site `http://www.csszengarden.com` is an illustration of the view that aesthetics and configurability of appearance are not mutually exclusive. The visitor is presented with a professionally designed web page. A control on the page allows the user to select a new style sheet that can completely change the site's colour scheme and layout. It *is* possible to create sites that are aesthetically pleasing without sacrificing usability.

The scale and sophistication of the possible recolouring is significantly greater when it has been anticipated and is incorporated in the initial design. Ad-hoc colour scheme changes should *not* be left as a challenge to the user of a web site.

2.2.3 Colour affects visual search and recognition

A number of studies have found that the inappropriate choice of colour can significantly affect the performance of users in such tasks as search and recognition. In comparing various text/background colour combinations on a web page, Ling and van Schaik (2002) found that reaction time and accuracy improved with increasing contrast. These results are corroborated by Ojanpaa and Nasanen (2003), who emphasise that luminance contrast is more important than chromatic contrast, especially for small text. Given these results, it is surprising that guidelines for the use of colour in the FAA (Federal Aviation Administration) had yet to be developed in 2006:

> *"At present, the Federal Aviation Administration has no requirement for how color should be used in ATC displays. While the advantages of color may be apparent, many display designs suggest that ATC technology developers have not used basic human factors and color principles to optimize the advantages of color use in complex scenes such as those in the ATC environment. In addition, technology developers create their own unique color schemes. The lack of consistency in color use can be confusing. Moreover, little attention has been devoted to the potential negative effects of color use on controllers' task performance."– Xing (2006)*

2.2.4 Colour, culture, emotions and commerce

In marketing, the colour of products and advertisements are an integral part of the overall image being presented and are critical to successfully marketing a product. Surprisingly, when the viewing time is very brief, black and white advertisements can out-perform those using colour (Meyers-Levy and Peracchio, 1995).

While, as noted in section 2.1.3, there is significant agreement on the emotional connotations of colour, there are also culture-specific colour scheme preferences and prohibitions that must be taken into account when selecting colour schemes by companies selling internationally (Morton, 2009a). Kondratova and Goldfarb (2006) surveyed sites from fifteen countries and identified palettes that appear to be globally acceptable, and those specific to particular countries. Kobayashi (1998) conducted a survey of cities and noted that not only are there culturally significant colours, but also colour palettes specific to particular regions within countries.

The risk of lost business caused by the use of inappropriate colour schemes and the complexity of selecting appropriate colours (e.g. to allow for gender, culture, age and desired impression) has led to new businesses as companies collect, collate and sell data resulting from web-based surveys:

> *"Data from Colorcom's international survey*[7] *will provide reliable information that a specific color is evocative and can subconsciously link the desired attributes to a product brand, an entity or image. Users can select the demographic group they are interested in and list the attributes (the communication or symbolism) they want the colors to achieve. For example, you could specify North American males, ages 24–35, (demographic group) and then request the colors that are associated with 'powerful.' As an alternative, if you're concerned that a color – such as green – might not communicate the desired attributes of 'dependability' (for example), a search for the color green would deliver a list of what green communicates to any demographic group." – Morton and Peterson (2007).*

The appeal of having such a database available to aid the automatic design of colour schemes is considerable, but the information is proprietary and its incorporation would add another dimension to the generation of a harmonious colour scheme. Before attempting to orient a colour scheme at a particular target group, it is first necessary to determine whether it is possible to create a colour scheme automatically. If this proves to be possible, the targeting could be added later as constraints on the usable colours.

An alternative to the survey approach of Morton is the analysis of existing web sites to derive a database of relevant features such as the "cultural markers" and "cultural attractors" of Kondratova et al. (2005).

Colour and the assessment of trustworthiness

A prerequisite to online sales is trust, and, online this must be established without using any of the usual cues normally available during human interaction. With the exception of sites with an established reputation (e.g. `http://www.amazon.com`), the assessment of whether a business is worthy of trust is based completely on the impression given by its web site. The colours used on the site have been found to play a significant role in a user's perception of the site's trustworthiness (Wang and Emurian, 2005; Karvonen, 2000; Wang and Emurian, 2005).

A topic related to trust is professionalism. Hall and Hanna (2004), in assessing the impact of text-background colour combinations, found *"a site that is viewed as*

[7] Surveying over 60,000 participants, as of May 2009.

readable is also viewed as professional", and that pleasing web site colour combinations, especially those based on blue, are more likely to result in sales.

Damhorst and Reed (1986) found that clothing colour had a significant effect on the perceived competence of job applicants, but this was gender-specific, applying only to male applicants.

2.2.5 Standards bodies and colour use

Standards bodies have not attempted to codify colour harmony. There are, however, usability guidelines such as ISO 3664:2000 and those from the W3 Consortium on web usability that include references to colour and readability, primarily related to minimum lightness contrast ratios for small and large text W3C (2007). The impact of colour on readability will be covered in section 2.9.8, but while it is obviously desirable for there to be a strong differentiation between foreground and background elements, whether for text or user controls, in practice, inadequate contrast on websites is widespread problem:

> *"a recent formal investigation conducted by the Disability Rights Commission (DRC) has brought attention to color accessibility. The study evaluated 1000 web sites for compliance with the World Wide Web Consortium (W3C) Web Content and Accessibility Guideline (WCAG) 1.0 checkpoints, defined by the Web Accessibility Initiative (WAI). Notable in the results is that 81% of the web sites tested failed to satisfy the most basic WAI compliance category. 'Inappropriate use of colors and poor contrast between content and background' accounts for 59 of the 585 usability problems" – Jefferson and Harvey (2006).*

It is clear, therefore, that colour, if used in an interface, must be used with care. Otherwise, rather than enhancing the interface, it can mislead the viewer with unwarranted or inappropriate emotional expectations, and possibly impair usability. In surveying existing sites, Nielsen and Hahir (2002) noted of one *"these color-coded site areas aren't obvious enough because the site uses so much colour everywhere else"*. The "fruit salad" look[8] – colour everywhere and no meaning – must be avoided. Colour may be used to convey information or for decoration, or both, but whatever the reason, it must be unambiguous to the viewer and appropriate to the context.

2.2.6 Neurophysiological factors affecting the use of colour

Processing within the human visual system can create visual effects or artifacts that alter the way objects are perceived. This can affect not only the perceived colour, but also the apparent spatial arrangement of objects and the perception of what is important and what is part of the background (*figure–ground* contrast (Birren, 1961; Stroebel et al., 1980)). Such perceptual effects can be caused by: abutting saturated complementary colours, which can cause unpleasant shimmering effects; simultaneous

[8] A term attributed to Marcus (1997) by Ling and van Schaik (2002).

contrast, which can alter the apparent colour based on nearby colours; and "warm–cool" colour contrasts, which can affect the apparent depth of objects. All of these can affect the appearance of coloured elements in an interface.

The effects of highly saturated complementary contrast

The human visual system has a highly developed sensitivity to the detection of edges. This results in several visual artifacts such as the emphasis of difference between abutting regions. This applies to both luminance and chromatic differences, and can result in strange effects, such as shimmering colours at the boundary of abutting complementary colours (figure 2.2); and the appearance of gradations in colour depending on the colour of adjacent areas, as illustrated in figure 2.3.

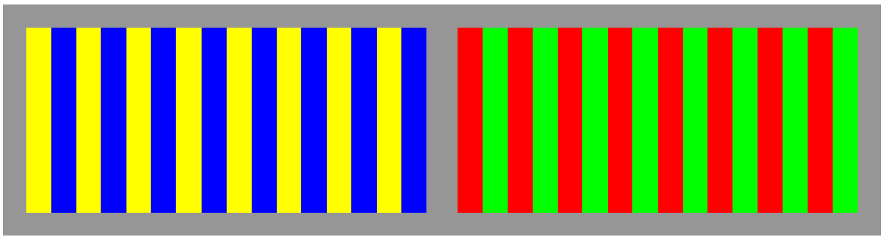


Figure 2.2: The adjacency of high saturation complementary colours can cause unpleasant shimmering effects at the boundary, an example of a visual artifact related to high saturation chromatic contrast.

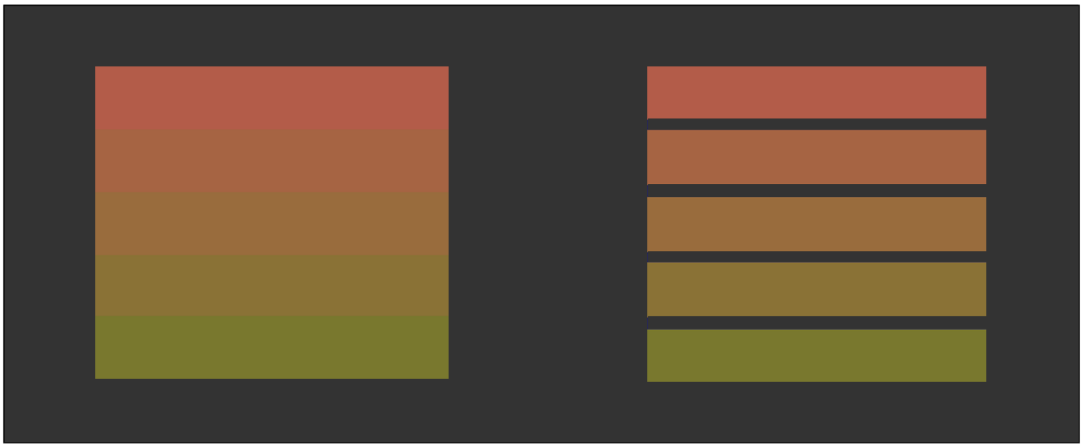


Figure 2.3: The gradations in colour apparent in the left-hand image are visual artifacts – each band is a single colour. As shown by the right-hand image, the gradation disappears if the regions are separated.

The effect of "visual weight" is discussed in detail by Tufte (1990, p60-61). He notes that excessive contrast introduces the appearance of unintentional and confusing negative space artifacts which he, following Albers (1969), calls the "1 + 1 = 3" effect (figure 2.4). This effect is another example of the perceptual ambiguity between what is important (the foreground) and what is part of the backdrop to the foreground objects (the background). While the effect is more likely to be problematic within diagrams, care should nevertheless be taken when colouring an interface, as with particular spacing and high contrast, it is possible induce the same effect.

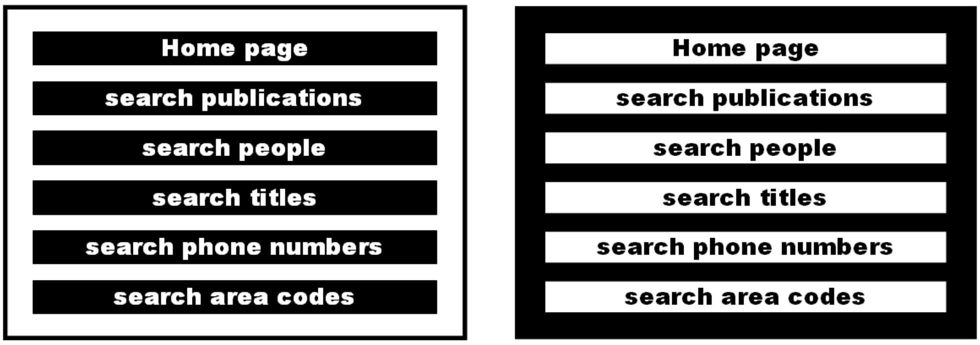


Figure 2.4: An example of the conflict between foreground, background and negative space in high contrast areas (based on an example by Tufte 1990). The approximately equal width of the text and the white spacers in the left image (black in the right) can cause the spacers to appear to be the foreground elements instead of the text.

Simultaneous contrast

The appearance of a colour depends markedly on the colours of adjacent areas[9], an effect described in the aptly entitled paper by Ishizaki (1995): *"What You See In Your Color Palette Isn't What You Get!"*. This effect, known as simultaneous contrast (or more technically, chromatic induction), has been known since ancient times. Aristotle wrote:

> *"bright dyes too show the effect of contrast. In woven and embroidered stuffs the appearance of colours is profoundly affected by their juxtaposition with one another (purple, for instance, appears different on white and on black wool)" – Aristotle (2010).*

A detailed study of the effects of simultaneous contrast of dyes were made by Chevreul (1855), who was employed as a chemist at the Gobelin dye works in Paris to correct an apparent inconsistency in the colour of dyes. He found that the dyes were not at fault, rather it was the appearance that changed. Dyed regions could take on a different colour appearance as the surrounding colours changed (Ball, 2001). This effect applies not only to adjacent coloured regions, but also to the abutment of chromatic and achromatic areas, and can lead to the appearance of colour where there is none: *"yellow demands purple"*[10] – von Goethe (1810).

The effect is explored in detail by artists (Itten, 1970a) and from a physiological viewpoint by Katz (1999) and Wesner and Shevell (1992). Research aimed at developing

[9] Interestingly, this effect is not restricted to humans, but also occurs in goldfish (Dorr and Neumeyer, 1997).

[10] A saturated colour on an achromatic background can result in the complement of the colour being seen due to after-image effects (i.e. seeing purple after looking at yellow).

models of simultaneous contrast effects is ongoing: (Miyahara et al., 1993; Barnes et al., 1999; Shepherd, 1999; Barnes and Shevell, 2002). The effects can also depend on colour and texture. Monnier and Shevell (2004) found that for some colours[11], patterned backgrounds could induce larger shifts than those of a uniform colour.

An interface is much more complex than the simple patches or concentric rings often used in the current psychophysical experiments. Considering more complex scenes, Brenner et al. (2003) state:

> *"If there are many surfaces, then not only the average luminance and chromaticity of the surfaces matters, but also the chromatic variability. It is not yet clear whether it makes any difference where the chromatic variability is within the scene, so we constructed stimuli in which the chromatic variability was restricted to certain regions. We found that it made very little difference where the chromatic variability was located. The extent to which the average colour of nearby surfaces influences the apparent colour of the target seems to depend on the average chromatic variability of the whole scene".*

This effect is not restricted to physical objects, but also applies to displays (Luo and Gao, 1995). The impact of simultaneous contrast in interface design is explored by Gurura et al. (2004) where the importance of the background in an interface is emphasised. The salience (visual prominence) of an interface item can be changed with no change to an item's colour, solely by changing the colour used as its background. Both Gurura et al. (2004) and Linnett et al. (1991) note that the background can significantly alter the perceived colour harmony: changing the background can make inharmonious items appear harmonious, and vice-versa. From this, we may infer that the background must be an integral part of the design of a colour scheme.

The effect is not restricted to interactions between objects within the scene, but also between the scene and the surround:

> *"Surround also plays an important part in the appearance of colour. The simultaneous contrast effect is well known for a single coloured patch on a coloured background, but applies also for a coloured border around a complex image. The effect of a dark border, for example, is to make the image appear lighter and less contrasty, whereas a light border makes the image appear darker and also less colourful." – MacDonald (1993)*

The use of computational models to quantify this effect has met with some success. Ishizaki (1997) has explored the possibility of automatically adjusting the colour of informational graphics to ensure that identically coloured items *appear* to have the same colour in spite of differently coloured backgrounds. The colour of an element is adjusted to compensate for the effect of its surrounding colour, or the average background colour when there is more than one. One aim of this work was to compensate for the effects

[11] These were colours that affected only one cone in the eye, that sensitive to short wavelength of light (the S cone). The cone sensitivities are discussed in section 2.3.

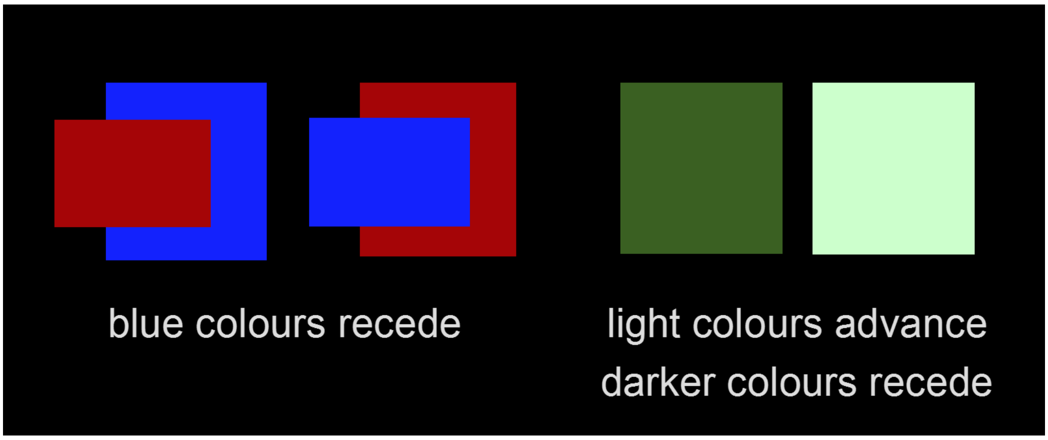


Figure 2.5: Left: warm–cool contrasts can cause tensions in figure–ground perception. A cool (blue) foreground object can make a warm background object appear to be wrapped around the cool object rather than behind it. Right: lighter colours appear closer to the viewer, even when of the same hue.

of simultaneous contrast to improve the visual cohesion of items meant to be seen as a group when seen on differently coloured backgrounds. The results indicate a significant, but not complete, compensation for the background colour changes.

It would seem premature to attempt to automatically incorporate or compensate for the effects of simultaneous contrast in the design of an automatic colouring system, as the effects of chromatic induction/simultaneous contrast are subtle, and it is clear from the literature that the models are, as yet, incomplete.

Depth perception is colour-dependent

Another visual artifact is the impression of apparent depth when one object is a warm (red, yellow or oranges) and the other a cool colour (blues or green). The warm-cool colour dimension appears to be one of the basic classifications of colour (Itten, 1970b).

Objects with cool colours often appear to recede or be further way than warm coloured objects Morton (2009c), due to the changes in focal depth caused by the different wavelengths as the light is refracted by the lens within the eye. This visual effect can be used by designers to draw objects into the foreground and give them prominence, but it can also lead to visual difficulties as it can cause objects that are intended to appear as further away (figure 2.5) to be (partially) promoted to the foreground. A related effect is that of lighter coloured objects appearing closer to the viewer.

2.2.7 Colour use in graphical user interfaces – summary

It can be seen that there is ample experimental evidence to justify the assertion that "the colours used in an interface do affect the viewer at an emotional and perceptual level" and it is therefore desirable to optimise the colour scheme of the user interface in order to engender the desired impression. There are however physiological factors that also affect the perception of colour. These will be discussed in the next section.

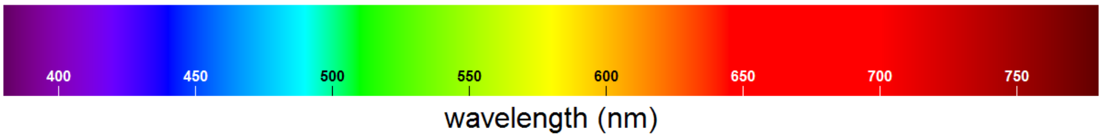


Figure 2.6: An indication of the correspondence between colour and wavelength in the range of human vision (380–750nm).

2.3 Human colour perception

Colour is a perceptual response to light between 380 and 750nm in wavelength, and the approximate correspondence between wavelength and perceived colour[12] is shown in figure 2.6[13]. As colour is a perceptual response, it is a subjective phenomenon.

Non-luminous objects must be illuminated to be seen. The source of illumination and the light reflected from an object will have a spectral power distribution (SPD), and the perceived colour will depend on the spectral power distribution. For example, if the peak of the distribution is towards the longer wavelengths, a colour from the red end of the range will be perceived, with the intensity of the colour being a function of the shape of the distribution.

To generate the perception of a colour, recreating the complete spectral power distribution of the light from an object would suffice. In the same viewing conditions, this would indeed recreate the perceived colour, but fortunately, this is not necessary. Colour is commonly represented by three numbers (e.g. the RGB colour values) – tristimulus values that are the basis for all digital encodings of colour. To understand why this is sufficient, and why the viewing conditions affect the perception of colour, it is necessary to understand, in outline, the human visual system.

The basis of human vision is the interpretation of an image projected onto the retina by a lens in the eye. The retina has two types of light sensing elements: rods and cones. Of the two, the rods will give outputs at illumination levels well below that required for the cones to function, but as the rods all have the same spectral response, their output can only indicate luminance, not colour. Therefore, at low light levels, colour vision is absent.

There are three types of cones, named the L, M and S cones after their ability to react to light at long, medium or short (LMS) wavelengths. The S cones have a peak sensitivity towards the blue end of the spectrum, the L cones towards the red end, and the M cones in between, as is shown in figure 2.7. The differing outputs from the three types of cones provide sufficient data for subsequent processing to enable colour perception.

The LMS cone peak sensitivities correspond approximately to the colours blue (S - 447nm), green (M – 540nm) and orange (L – 577nm). It is not immediately apparent

[12] The purple colours are not present in the illustration as purples do not correspond to a single wavelength. Purple is seen when both short and long wavelengths (blues and red) are present simultaneously.

[13] Created using a derivative of Earl Glynn's "Spectra" program from `http://www.efg2.com/Lab/ScienceAndEngineering/Spectra.htm` (accessed May 21, 2010) which credits Dan Bruton's Color Science page: `http://www.physics.sfasu.edu/astro/color.html` (accessed May 21, 2010).

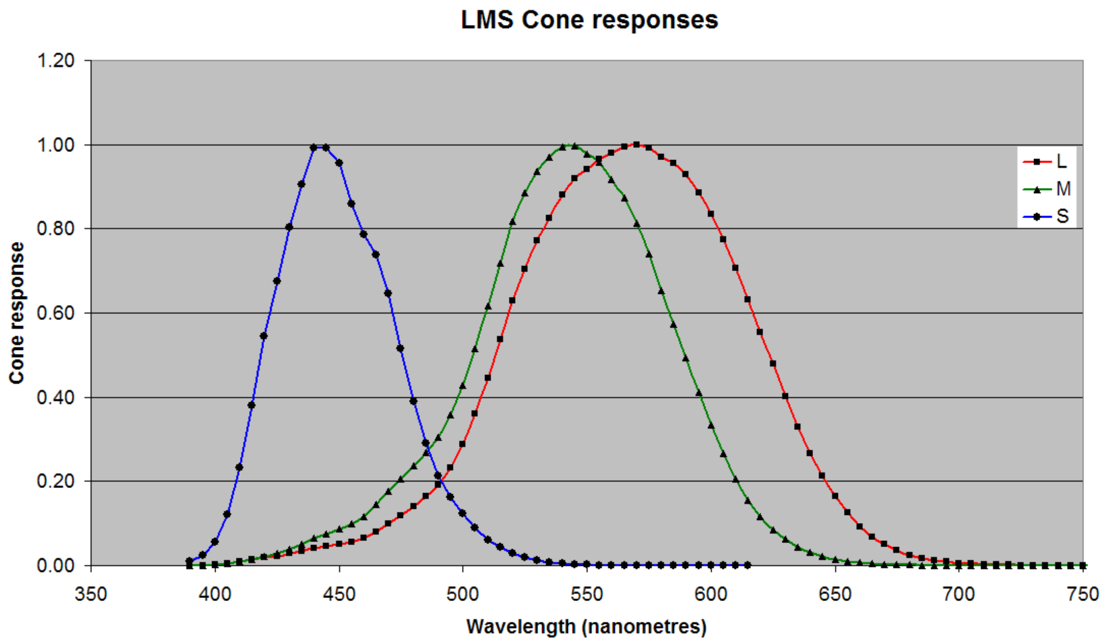


Figure 2.7: LMS cone responses to the varying wavelengths of light. The shorter (lower) wavelengths correspond to blues, the longer to reds. Plotted from Stockman and Sharpe (2000) 2° cone fundamentals.

how to reconcile the three cone signals with the four perceptually important colours: red, blue, green and yellow.

2.3.1 The derivation of perceptually important primaries

In 1810, Goethe published a wide-ranging treatise that is commonly known as a "Theory of Colours"[14] on his thoughts and observations about colours and its emotional connotations (von Goethe and Matthaei, 1971; von Goethe, 2006). Included was his "chromatic circle" – an arrangement of colours with the hues arranged perceptually, based on the observation of after-images[15]:

> *"the colours diametrically opposed to each other in this diagram are those which reciprocally evoke each other in the eye. Thus, yellow demands purple; orange, blue; red, green; and vice versa: thus again all intermediate*

[14] "Theory of Colours" is the name given as the title by Charles Westlake in his English translation of Goethe's work. This appears not to be correct. Gimbel (1993) gives the correct title as *The Teachings of Colour*. MacEvoy (2008), having retranslated the work, also comments on the erroneous title and notes other shortcomings in the translation, such as its incompleteness, with whole sections missing from the English translation. MacEvoy: *"Unfortunately, Goethe's ambitious project has been rendered incoherent both by the deleted sections and by the English translation title: Farbenlehre simply means 'chromatics,' with no 'theory' implied (just as Sprachlehre means 'grammar' and not 'theory of speech'). Given Goethe's sensitivity to language, it is not irrelevant to note that the root meaning of lehre is 'lesson,' 'teaching' or 'learning from experience' "* – MacEvoy (2008). For details on the omitted sections and a critique, see MacEvoy's extensive web site: `http://www.handprint.com` (accessed May 21, 2010).

[15] An after-image is the image seen after fixating on a highly saturated colour for 30–60 seconds and then looking at a white area. The complementary colour will be seen.

> *graduations reciprocally evoke each other" – von Goethe (2006), ¶50*[16].

There had been several prior three-dimensional arrangements of colour, but Goethe gives notes the interpolation of the four primary colours and the separateness of blacks, whites and greys:

> *"...yellow, blue, red, green. They represent the most general idea of colour to the imagination, without reference to any very specific modification.*
> *If we were to add two other qualifying terms to each of these four, as thus — red-yellow, and yellow-red, red-blue and blue-red, yellow-green and green-yellow, blue-green and green-blue, we should express the gradations of the chromatic circle with sufficient distinctness; and if we were to add the designations of light and dark, and again define, in some measure, the degree of purity or its opposite by the monosyllables black, white, grey, brown, we should have a tolerably sufficient range of expressions to describe ordinary appearances" – von Goethe (2006), ¶610 & 611.*

In this, Goethe produces the first perceptually derived colour arrangement of a hue circle (shown on p41), and notes that black, white and grey (and incorrectly, brown) are different from the other colours.

In 1892, Ewald Hering published his "Opponent Process Colour Theory" (Hering, 1964) in which he theorised the existence of four *perceptual* primary colours arranged as two complementary axes – a red–green axis and a blue–yellow axis – together with a lightness axis. Hering based this theory on observations that, in language, certain colour pairs are never used together when describing colours. There isn't a blueish-yellow or a greenish-red, suggesting something special about those pairs. The study of after-images makes the opponency of these pairs quite evident. He also noted that colour vision deficiencies often affected either red–green or blue–yellow discrimination. Hering's theory was later validated experimentally by Hurvich and Jameson (1957).

Hering theorised that there were differing receptors in the eye responsive to light-dark, red-green, and blue-yellow differences. At the time, this supposition was not well received, but while not strictly accurate, it is now known to be substantially correct: there are three sensing elements, and the axes are as he anticipated, but the opponent signals are derived, not primary.

As illustrated in figure 2.8, the opponent colour signals and a luminance signal are calculated from the outputs of the LMS cones by the ganglia before being sent to the optic nerve and the brain (Fairchild, 1998). This recoding of the information by the ganglia transforms the LMS signals into the red–green, blue–yellow and light–dark percepts posited by Hering. Physiologically, addition and subtraction are effected as excitation and inhibition of neurons: for a detailed discussion of neural opponent processing, see Abramov and Gordon (1994).

[16] paragraph numbers (e.g. ¶50) indicate the location in the translated or annotated text.

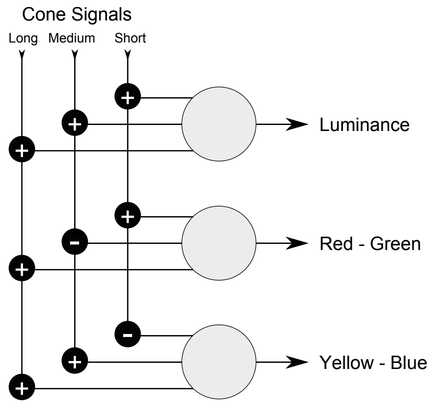


Figure 2.8: The outputs from the long, medium and short sensitive cones in the eye are combined by the ganglia to derive luminance and opponent colour signals, redrawn from Fairchild (1998).

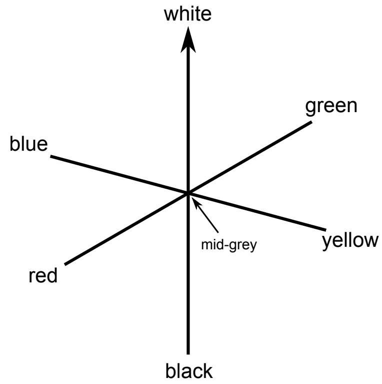


Figure 2.9: The dimensions resulting from the opponent processing: one for lightness, two for colour.

2.3.2 Colour vision deficiencies (CVD)

There are individuals with deficiencies in their colour vision who, for the most part, function perfectly well within society. The widely-used term *colour blindness* is a misnomer: monochromatism, the complete inability to distinguish colours, is extremely rare, affecting around 0.005% of the population (Fairchild, 1998). Those usually classed as colour blind have either altered colour perception (anomalous trichromacy) or the inability to distinguish between certain colours (dichromacy); the frequency of occurrence is shown in table 2.1.

Medical Term	Type	Male	Female
Monochromacy		0.003%	0.002%
Dichromacy	Protanopia	1%	0.02%
	Deuteranopia	1.1%	0.01%
	Tritanopia	0.002%	0.001%
Anomalous trichromacy	Protanomaly	1%	0.02%
	Deuteranomaly	4.9%	0.38%
	Tritanomaly	∼0	∼0
	Total	∼8.0%	∼0.4%

Table 2.1: The incidence of colour vision deficiencies in Western races, derived from Fairchild (1998).

Anomalous trichromacy is most common, affecting around 6.3% of the population (5.9% male, 0.4% female). In anomalous trichromats, the response curves of the LMS cones differ from the norm, resulting in altered colour perception. The effects can be quite subtle and the impairment may only be detected during a colour vision test. The reason for the predominance of those with CVD being male is the adjacency of red–green photo-pigment genes in the X chromosome. In the case of mixing errors, females have another copy whereas males do not (Manniesing, 2003).

Dichromacy is less common, affecting ∼2.1% of the population, and occurs when one of the colour receptors is either missing or dysfunctional. This results in the viewer losing the ability to distinguish a range of colours. The most common forms are: protanopia, the loss of the red receptors, which affects ∼1% of the male population; and deuteranopia, the loss of green receptors, which also affects ∼1% of the male population. Both result in the loss of red–green–yellow discrimination. The percentage of females affected for both forms is small (0.03%). Impairment of the blue cones (tritanopia) is very rare, affecting less than ∼0.003% of the population.

Overall, the most common colour vision deficiencies cause altered colour perception and are primarily restricted to males, with around 8% of the male population and 0.4% of the female population being affected. The deficiencies alter the perception of colour schemes, but do not impair lightness discrimination. If a subgroup of viewers is being addressed, the design of colour schemes could be tailored to limit the effect of a specific impairment. However, omitting hues that cause difficulty for those with the most common impairments would seriously limit the colours that could be used. Therefore, the system being outlined in this research does not specifically address colour deficient viewers. However, noting that lightness perception is unaffected, it may, by ensuring lightness contrast, create schemes that are usable by colour deficient viewers without limiting the hues used.

2.4 CIE colour spaces

Research into human vision resulted in the publication of the luminous efficiency curve (CIE, 1926) by the international standards body, the CIE (Commission Internationale de l'Eclairage – the International Commission on Illumination). The luminous efficiency curve specifies the sensitivity of the eye to different wavelengths (perceived as different colours). Several years later, based on the work by Wright (1929) and Guild (1931), the CIE defined two colour spaces, the CIERGB colour space and the CIEXYZ colour space, which give a measure of the necessary amount of three defined primary colours to produce the visual impression of all perceivable colours by the *CIE Standard Observer*.

To recreate the impression of a colour, it is enough to stimulate the LMS cones so that they produce the same outputs as the original colour; the spectral power distribution of the light stimulus does not have to be identical. Recreating a colour impression without recreating the original spectrum is known as metameric colour matching and is the basis of all common display and image capture devices. In a metameric match, two differing spectral power distributions result in the same perceived colour. A metameric match will only be correct under the same lighting conditions, which is why two objects that appear to match in colour with one source of illumination may appear to be different colours under another.

A colour sensation C can created as a mixture of three primaries:

$$C = rR + gG + bB \tag{2.1}$$

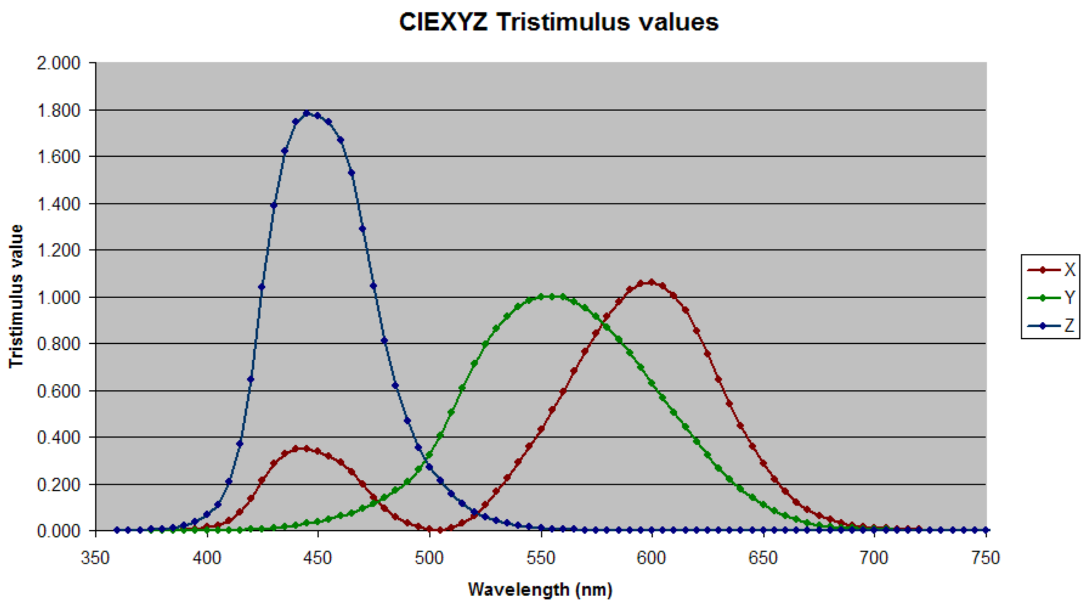


Figure 2.10: The CIEXYZ colour matching functions that indicate the contribution of three imaginary primaries to create the appearance of colours for the range of human colour vision.

where r, g, b are scaling factors for the primaries R, G, B. The scale factors (r, g, b) are known as the *tristimulus values*. With known primaries (R, G and B), using just the three tristimulus value, is possible to recreate the impression of a colour, as long as the lighting conditions are identical.

The CIERGB space is a definition of the range of visible colours in terms of an additive mixture of three primaries: red, 700nm; green, 546.1nm and blue, 435.6nm. In colour matching experiments, a specified primary colour (the target) was matched by varying the ratio of light from three coloured sources, so the blended source light appeared to be the same colour as the target. It was found that some of the more saturated target colours could not be matched with any combination of the sources, and a match could only be found by shining one of the source colours onto the target being matched (rather than blending with the two other sources). Illuminating the target with a source colour desaturates the target colour and enables it to be matched by the two remaining primaries. These situations are reflected in the tristimulus value for that primary being negative. This was not ideal, as a significant amount of calculation is involved when transforming colours, and as this work was conducted prior to the availability of computers, it was much more convenient if the values were all positive. This was one of the considerations in the design of the CIEXYZ colour space.

2.4.1 The CIEXYZ colour space

The CIEXYZ space is a transformation of the values from the colour matching experiments designed to give tristimulus values (X,Y,Z) that are always positive, and to ensure that one set of tristimulus values (Y) would correspond to the photopic luminous efficiency (lightness sensitivity). The CIEXYZ tristimulus values are shown in figure 2.10. The CIEXYZ colour space makes use of primaries that are not physically realisable. These imaginary primaries therefore cannot be used as the basis of any display device, but the CIEXYZ space is useful as it can be used as a device-independent definition of the humanly perceivable colours, and can therefore be used as a intermediary between any two colour spaces. For details on the derivation of the CIE colour spaces, see Sharma and Trussell (1997) and Fairchild (1998). Poynton (2003) approaches the derivation from the unusual perspective of spectral power distributions (SPD) and filters of differing bandwidths.

2.4.2 The CIE chromaticity diagram

At any wavelength, a set of XYZ values can be normalised so that the total magnitude is 1. This removes the magnitude (the lightness) and leaves the ratios, x, y and z:

$$x = \frac{X}{X+Y+Z} \tag{2.2}$$

$$y = \frac{Y}{X+Y+Z} \tag{2.3}$$

$$z = \frac{Z}{X+Y+Z} \tag{2.4}$$

One term is redundant as $z = 1-x-y$, and so any spectral colour can thus be specified by x and y. These values are called *chromaticity coordinates.* The transformation preserves the ratios between x, y and z, but their magnitudes (corresponding to lightness) are lost, leaving only the hue and saturation – the chromaticity information.

The plot of x and y for the visible spectrum is a projection, and is known as the CIE Chromaticity Diagram (figure 2.11). The curved edge of the shark-fin shaped area is known as the spectral locus, and corresponds to the fully saturated colours perceivable by a *typical* human observer. Points inside the curved area correspond to differing ratios of X,Y,Z excitation.

An important coordinate is the point where all three of x, y and z are equal: $x = y = z = \frac{1}{3}$, which corresponds to white (point E on the diagram). Moving away from the white point towards the spectral locus (the edge of the fin-shaped region) corresponds to increasingly saturated colours of the same hue. The colours seen as purple do not correspond to a single wavelength, but are blends of red and blue. On the chromaticity diagram, the purples appear along the flat line joining the long and short wavelengths.

If the chromaticities of any three "primaries" (red, green and blue are shown in fig. 2.11) are plotted on the diagram, the enclosed triangular area indicates the *gamut* – the range of representable colours that may be obtained by mixing the ratio of the three primaries.

The chromaticity diagram is a useful visualisation of the range of possible colours. If two sets of chromaticities are plotted, the difference between the triangles is an *indication* of the differences between the gamuts. To accurately gauge gamut differences, a three dimensional plot is required (Fairchild, 1998). The chromaticity diagram is

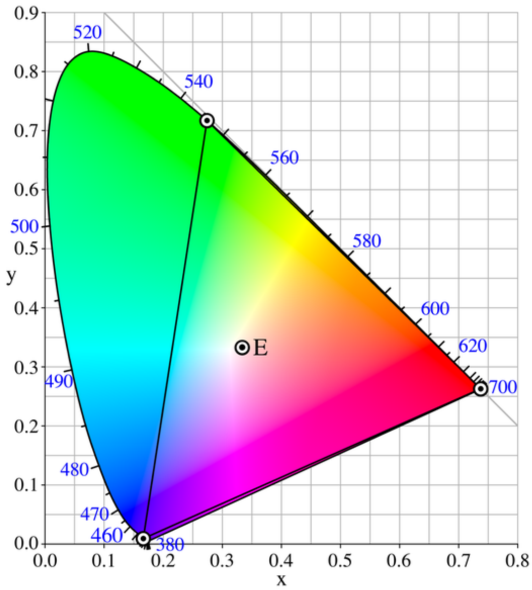


Figure 2.11: The CIE Chromaticity diagram

widely used and often coloured, but this colouring is an approximation, as the colours are limited by the gamut of the display or printer used to render the diagram.

2.4.3 The white point

To be perceived as white, it is not necessary for the chromaticity to be exactly at the equal energy point E ($x = y = \frac{1}{3}$). After a period of adaptation (to be discussed in section 2.5.1), there is significant latitude in the chromaticities that will be perceived as white.

To define which colour will be "white", chromaticity coordinates could be used, but it is more common to specify a *correlated colour temperature*. As a black body is heated, it begins to glow first red, then orange, yellow and a bluish-white. An unambiguous definition of colour can be achieved by specifying a temperature at which the temperature of a black body radiator most closely matches a particular colour, in this case a shade of white. This is called the *correlated colour temperature*. By specifying a colour temperature, an unambiguous definition of colour is achieved.

For example, a monitor with a 5000K white point will have the balance of red, green and blue adjusted so that when all three are fully excited, the resulting white has the same apparent colour as a 5000K black body radiator, which would correspond to a warm white.

2.4.4 Standard Illuminants

The spectral composition of the illumination affects the perceived colours. This has implications for colour reproducability and can be especially important in an industrial setting. To define typical lighting conditions, the CIE have specified "standard illuminants" and defined their spectral distribution. The illuminants relevant here are those that relate to typical working conditions for computer users: *CIE Illuminant A* – 2856K, incandescent lighting; *D50* – 5000K, warm, early or late horizon daylight; and *D65* – 6504K, daylight at noon (Hunter Associates Labs, 2008). These illuminants allow the standardisation of viewing conditions so that colours, including those on a display, may be seen as intended, but as discussed in the next section, such standardisation is rarely necessary.

2.5 Adaptation of human colour vision to the viewing conditions

There are many differing ways of transforming the three tristimulus values representing a colour, with each transformation resulting in a different colour space. Whatever transform is used, the colour value is represented by three numbers. These tristimulus values are the magnitudes of three primaries that have a defined chromaticity, measured in a particular context, with illumination of a known level and spectral balance. If any of these differ, so will the colour appearance. Whether or not the resultant difference in colour appearance is important to the viewer depends on the application. In some applications, an identical colour appearance is necessary; in others, an approximately correct, but pleasing, appearance is sufficient.

Professional graphic designers working with colour images require an on-screen image to accurately reflect the colours that will result when the image is printed. For this to be possible, the characteristics of displayed colours must be precisely defined. For identical *colour appearance*, the on-screen image must not be desaturated or altered by reflected ambient light. It is therefore necessary to control both the level and the colour temperature of the ambient illumination, as is evident here:

> *"The Adobe RGB (1998) color image encoding is defined as an encoding of the color appearance of an image that is being displayed on a reference color monitor in a reference viewing environment.*
> *NOTE The intended color appearance can be reproduced exactly on a physical device in an actual viewing environment, only when the actual viewing environment exactly matches the reference viewing environment... When measured, with the monitor turned off, at the monitor faceplate, the ambient illumination level shall be equal to 32 lx. ... The ambient illumination shall have the same chromaticity as the white point of the display." – Adobe Systems (2005)*

Few computer users require such precision in the reproduction of colour. As long as the colours displayed are sufficiently bright and colourful, and approximately correct, not only will they will be happy with the image, they won't even realise the colours are not "correct".

Empirical evidence suggests that strict control of viewing conditions is not too important for normal day-to-day use of on-screen colour. Colour schemes – sets of colour intended to be seen together – are developed by users all over the world and

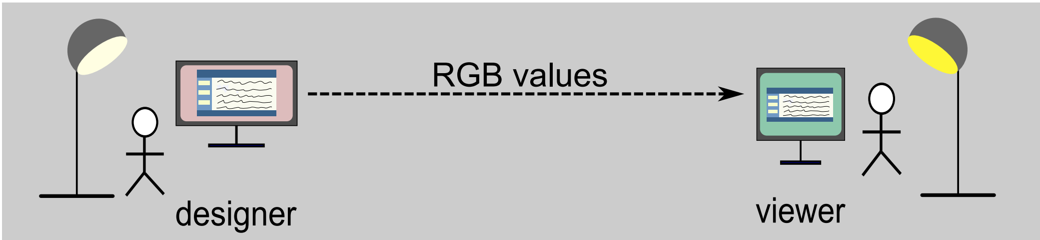


Figure 2.12: The sources of variability in colour scheme creation include both the designer's and the viewer's preferences, equipment and viewing conditions. Nevertheless, colour schemes can be created and disseminated, with a large degree of agreement as to their visual appeal, using only triplets of RGB values.

seen by others, in different countries using different equipment and different viewing conditions (fig. 2.12). Yet, by and large, there is widespread agreement on the visual appeal of harmonious colours, as witnessed by the popularity of particular themes and colour schemes on the previously mentioned sites for theming web applications (sec. 2.2).

Typical users, apart from adjusting the brightness and contrast (often to the maximum setting), adjust neither the colour settings on their display nor within their computer. The colours displayed by their monitors are therefore only approximately correct. This is true even on displays that have been calibrated and for which a colour correction profile exists. The viewing conditions must be tightly controlled for correct colour appearance, and this occurs rarely.

2.5.1 Colour constancy and adaptation

Adaptation in the human visual system is the reason for the non-criticality of the displayed colours. The user still "see" the same colours, even if the reflected spectra of physical objects has changed significantly as the spectrum of the illumination changes. Unless the change in the spectrum of illuminating light is extreme, a human observer is unlikely to notice.

The human visual system has several mechanisms to stabilise the apparent colour even though the wavelengths from a scene can vary widely. Part of this is the ability to "discount the illuminant" – effectively renormalising the visual system so that, after a period of adaptation, the colours appear as expected. Although full adaptation appears to be limited to illumination in the range 5000K-6500K (Poynton, 2003), incandescent lighting (~2800K) appears acceptably close. It has a slight, but not objectionable, yellowish tint. The spectrum of the light from the setting sun is very different from that at midday, but the colours of many objects do not appear to change. Another cognitive mechanism appears to preserve the expected colour for objects with well defined colours and work "from memory", rather than what is actually being seen:

> *"in everyday life we are accustomed to thinking of most colors as not changing at all. This is in large part due to the tendency to remember colors rather than look at them closely." – Evans (1943), cited in Fairchild (1998).*

Hubel (1999) makes the interesting point that although the spectrum of light at sunset almost exactly matches incandescent lighting, the visual impression is quite different. The impression of golden light at sunset does not happen when viewing objects lit with incandescent lighting. In a scene lit with incandescent lighting, all the illumination, both direct and reflected, is derived from the same source, whereas sunrise and sunset scenes have two lighting sources: the golden sunlight, and the sky. The sky has a colour temperature in excess of 10,000K and provides much of the shadow illumination. As the sun sets, the low angle shadows form a larger fraction of the total scene, providing a reference other than objects illuminated by the setting sun. This prevents complete adaptation and gives rise to the golden evening scenes.

Interestingly, this adaptation, if made apparent, can be disturbing. In a house lit with incandescent lamps, the placement of a single 6500K lamp in a bedroom was found to be particularly disturbing to the room's occupant[17]. Having acclimatised to the incandescent lighting in the remainder of the house, they walked into their bedroom and turned on the light: the room appeared to be lit with a ghastly shade of bright blue. Yet the same "daylight" colour temperature lighting is widely used in business and workshop environments without any suggestion of a blueish tint, due to the uniform lighting (all 6500K) and the user's adaptation.

In the same way, colours on a display can vary over a wide range and yet be perceived by users as perfectly acceptable. Graphic designers typically have the white point of their displays set at 5000K. This gives a white that appears yellow when compared with the white on a display set to a 6500K white point, and the white from a 9300K setting (another common option (Zuffi et al., 2007)) will appear to have a bluish tinge. Yet, after using any one of the three for a few minutes, its white will appear pure white, irrespective of any initial impression. This adaptation enables users to feel comfortable using monitors with widely-varying white points, even though, in an absolute sense, the colours are not accurate. This adaptation is also noticeable in the variety of colour settings found to be acceptable on television sets.

The system envisioned in this thesis is intended to produce harmonious colour schemes with a pleasing appearance for users in a *typical* viewing environment. Typical users do not have calibrated systems nor reference viewing environments. It is therefore essential that any colour schemes created using such a system are both usable and aesthetically pleasing when viewed in a typical non-reference environment.

2.6 Colour display systems

The use of red, green and blue (RGB) primaries is the basis of all common colour displays. The chromaticities of the primaries will determine the gamut – the range of displayable colours. There are many different RGB colour spaces (Süsstrunk et al., 1999), differing in the chromaticities of the primaries, the white point (which variation of "white" results when the maximum output is being generated by all of R, G and B)

[17] Not the author: this impromptu experiment arose as a result of the author purchasing "daylight" compact fluorescent lamps and using one to replace a failed incandescent lamp in a bedroom of another member of the household.

and the gamma (a measure of the non-linearity of the relationship between excitation and luminosity). All of these can vary from one RGB colour space and one display technology to another, so the specification of RGB values such as RGB = (250, 192, 110) only loosely specifies the displayed colour.

To more precisely define the displayed colour, the ambient lighting conditions, the overall gamma, and the characteristics of the display must be known. If these can be (even partially) standardised, there will be a closer match between the intended and rendered colours.

The visual system's tolerance of white point differences is fortunate, as in most environments, controlling the illumination is impractical. In contrast, significant differences in the chromaticities of the display and the system gamma can cause noticeable and objectional errors, especially in the mid-tones. A pragmatic agreement between the manufacturers of display hardware and those creating operating systems was instrumental in reducing this source of error.

As the cost of memory and high-resolution display technology decreased, the use of colour displays capable of displaying millions of simultaneous colours became commonplace. With systems capable of displaying high resolution photographs and the growing availability of affordable digital cameras and scanners, the lack of standardisation in colour representation became problematic: users expected, quite reasonably, that the colours in a photograph or scan would appear to be the same when displayed on the screen as when printed. Without standardising the colour space being used or the mapping between different known colour spaces, there is insufficient information for this to be possible[18]. To address this, Hewlett Packard and Microsoft proposed the use of a standard known as sRGB as a common standard for consumer-level colour representation. The primaries used by sRGB are the same as the ITU-R BT.709-5 HDTV (high definition television) standard, with a 6500K white point and a gamma very close to 2.2 (Anderson et al., 1995; Stokes et al., 1996).

The sRGB colour space has been widely adopted by manufacturers and has simplified the representation of colour in the domestic and small-business market. The standard specifies the primaries, the white point, the gamma, and the viewing conditions. If devices adhere to the chromaticity, white point and gamma standards, even without "correct" viewing conditions (due to adaptation in the human visual system), generally acceptable colour rendering is achieved. This allows users to create documents using uncalibrated systems in whatever lighting is convenient or comfortable, and other users to view these documents with similarly ill-specified equipment and uncontrolled viewing conditions, and it to be rare for there to be anything obviously wrong with the colours. The standard was not intended for the professional market, where both absolute colour matching and the largest possible gamut are expected. These require very strict constraints on viewing conditions and calibrated equipment, neither of which can be realistically expected from domestic and small business users.

Even though the gamut of the sRGB colour space is a subset of those colours representable by cameras, scanners, displays and printers, sRGB has been very successful.

[18] This brief overview ignores problems due to gamut mismatches.

A user with a digital camera producing images to the sRGB standard can display these images on their computer and have them printed by a photo-lab to their satisfaction, without being aware of any colour management taking place.

Almost without exception, all current non-professional computer displays default to an sRGB rendering, although the white point may be set higher than 6500K standard. This has the effect of making the images appear brighter and more colourful (the Hunt effect Fairchild (1998)). Most users rarely adjust their monitors, and the fact that images from Internet sites around the world are rendered in acceptable colour testifies to the success of the combination of the sRGB standard and the human ability to adapt to widely-varying white points. This level of standardisation is useful when algorithmically creating colour schemes. If the colours within a schemes are created to the sRGB standard, it is likely that the colours seen will be close to those intended. Should more precisely controlled colours be desired, it is possible to use colour management.

2.6.1 Colour management

The use of a standard such as sRGB significantly reduces the error that might be expected due to the display technology. However, even with colour values corresponding to the sRGB standard and sRGB-specified equipment, manufacturing tolerances and device adjustments result in the displayed colour differing (in an absolute sense) from that intended. To ensure the display is rendering correct colours, it is possible to characterise the display by displaying a set of known colours, measuring these colours to find out what colour is actually being displayed and creating a mapping table to correct for any discrepancies. A standardised method of representing these mappings has been defined by the International Color Consortium (ICC) (International Color Consortium, 2004).

The mappings (Color Profiles) define a transform between a source or destination colour space and a standardised intermediate colour space is called the *Profile Connection Space* (PCS). The CIEXYZ colour space is often used as the PCS. For colour spaces s_1 and s_2, if transforms exist between any both colour spaces s_1 and CIEXYZ, and s_2 and CIEXYZ, then as long as the colour being transformed is within gamut, colourimetrically correct transforms are possible between s_1 and s_2.

By measuring the chromaticities of the display, it is possible to determine its gamut and produce transforms to correctly render colours specified in CIEXYZ coordinates (as long as the colour is within gamut). There are, however, many sources of colour data (cameras, scanners, etc.) and many destination devices, and their gamuts differ. When mapping from one device to another, it is useful to have an intermediate colour space capable of representing all visible colours. This limits the number of transforms required for m source colour spaces and n destination colour spaces to $m + n$ rather than $m \times n$.

ICC mappings are useful when it is necessary to ensure that stored colour values refer to a specific colour in an absolute, rather than a device-dependent, manner. It enables a display (or printer) to reflect the colours of the originally encoded image. Nielsen and Stokes (1998) detail the derivation and transformations embodied in an sRGB ICC profile using CIEXYZ as the profile connection space. ICC profiles, while desirable, are not widely used.

A harmonious colour scheme can be represented by sets of sRGB values. In the unlikely event that the user is using colour management and controlled viewing conditions, the displayed colours will be very close to those intended. However, for the more normal case – no colour management and arbitrary lighting, due to adaptation and the standardisation of display technology[19] – we can be confident that the colours and balance of the colour scheme are seen as intended.

None of this guarantees that the viewer will think the scheme is harmonious. There are personal preferences to be allowed for, but if the colour scheme does not appeal, this will not be because of faults in the technology.

[19] An sRGB display is assumed.

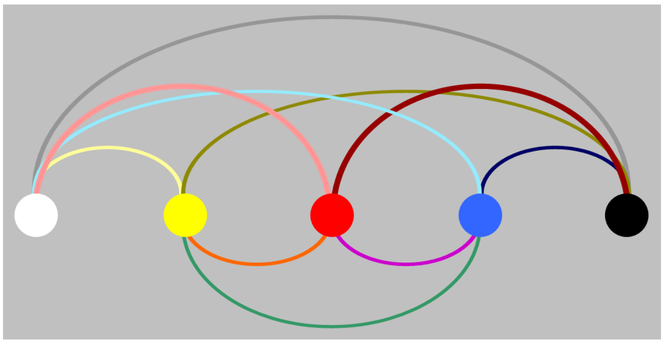


Figure 2.13: The ordering and relationship of colours of Franciscus Aguilonius giving prominence to red, blue and yellow, and ordered by the lightness of the colours (from 1613), based on a diagram in Norman (1990).

2.7 Colour spaces useful in colour scheme design

It has been said that the problem with standards is that there are so many to choose from. This is especially true of colour. In Kuehni's extensive study "Color Space and its Divisions" (Kuehni, 2003), there are thirty subsections in the historical chapter alone each describing at least one colour arrangement. The CIE-based colour spaces add to this number. The colour spaces can be divided into two primary categories: those whose organisation is based on physics, and whose organisation is based on perception.

As noted earlier, colour space is three dimensional and the space may be organised in many different ways, some of which are more useful than others in deriving harmonious colour schemes.

There are many theories about the derivation of harmonious schemes, but two principles seem fundamental: the use of complementary colours and the importance of *apparent* visual order (Westland et al., 2007). Complementary colours are those that, when mixed[20], blend to grey.

Before expanding on the subject of colour scheme design, it is necessary to introduce colour wheels and more colour spaces. Once these colour spaces and their properties (such as perceptual uniformity) have been added to the vocabulary, it is possible to detail how these colour spaces can be used in the design of harmonious colour schemes.

The circular arrangement of colours is now universal, but this has not always been the case. For a very long time, arrangements of colours were linear, frequently following the ideas of Aristotle, who ordered the colours from light to dark (Gavel, 1979). An example derived from the work of Aguilonius (a natural philosopher, 1567-1617) is shown in figure 2.13.

Newton was interested in the physical properties of light, in particular refrangeability (refraction). In his famous experiment of splitting white light into spectral colours

[20] The mixing may be additive or subtractive. Additive colour mixtures are those created by adding light, such as colour displays, dithered regions or visual mixing using spun disks. If all the primaries are present, the result is white. Subtractive mixing absorbs light, and having all the primaries (CMY – cyan, magenta and yellow) present results in (ideally) black.

and then the insight of wrapping the spectrum into a circle, Newton linked physics and the artistic arrangement of colour on a colour wheel. Newton was interested in the physical properties of light and the arrangement of the colours in Newton's colour wheel (fig. 2.14) is based on the position and proportion of colours in the spectrum, not on perception or any artistically oriented motive. For more detail on the construction and symbols on the disk, see Kuehni (2003, p43) and MacEvoy (2009b).

Newton's physically-oriented arrangement had a fundamentally different rationale for organisation of colours from the perceptual arrangement of Goethe (fig. 2.15) who was interested the psychological and emotive aspects of colour. Matthaei emphasises this distinction in the preface to his edited arrangement of Goethe's Colour Theory: *"It is particularly important to recognise that Newton and Goethe followed totally different aims in their research. While Newton attempted to analyse the nature of light, Goethe applied himself to the phenomenon of color. He wanted 'to marvel at the color's occurrences and meanings, to admire and, if possible, to uncover color's secrets'." – von Goethe and Matthaei (1971) quoted in Norman (1990).*

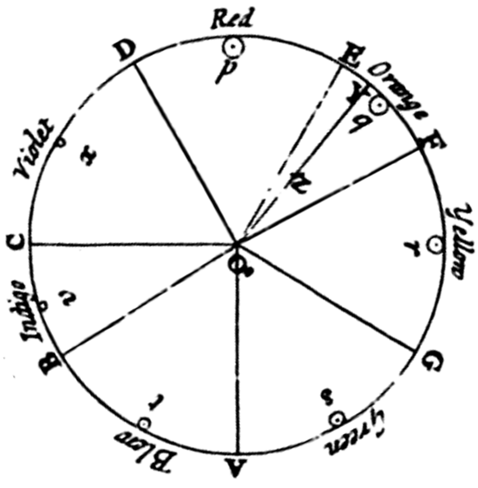


Figure 2.14: Newton's sketch of his colour wheel. The size of each segment is related to its width in the spectrum.

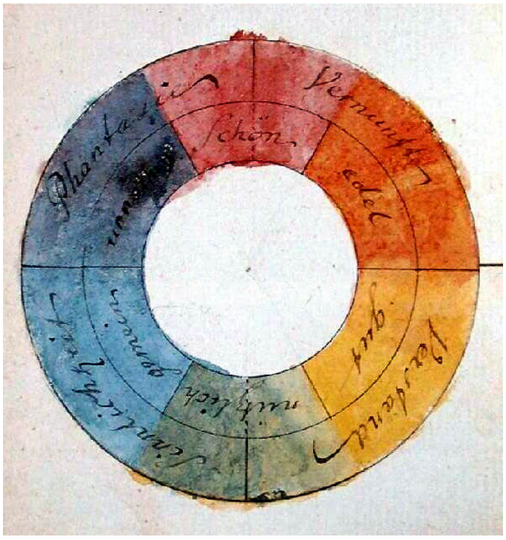


Figure 2.15: Goethe's colour wheel with each colour positioned opposite its perceptual complement.

In his encyclopedic treatise "Color and Meaning", Gage (2000) discusses the differing approaches to colour harmony of those following both Goethe and Newton. It is clear that Newton was familiar with the concept of complementary colours (MacEvoy, 2009b), but his colour circle, being based on physical measurements not perceptual ones, does not (intentionally) reflect this. Although Newton's colour circle was not intended for use in an artistic context, it was nevertheless enthusiastically adopted as a means of deriving harmonious colour schemes. Gage quotes Antonio Conti, writing of J.C. Le Blon, an artist and contemporary of Newton:

> *"the Newtonian theory of colours has given many [painters] the opportunity of determining their compositions by the mechanical rule of the centres of*

> *gravity [a concept deriving from Newton's circular mixture diagram]. That German painter who prints pictures [Le Blon] derived his secret from this source.... I met [him] at The Hague, and he assured me that following Newton's principles of the immutability and unequal refrangibility*[21] *and reflexibility of the rays of light, he had established the degrees of strength and weakness which colours require to be harmonized..." – Gage (2000, p139).* No date is given but it is likely to have been around 1720.

The colour wheels of Newton and Goethe position colours in different places (figures 2.14 & 2.15), which would tend to imply that there is some latitude in the selection of complementary colours.

Quoting Gage (p142): *"by about 1800, both scientists and painters had come to believe that the simplest form of colour-harmony was in the juxtaposition of complementaries"*. It is therefore clear that by 1800, the fundamental ideas that would permeate discussions of codifying colour harmony for the next two hundred years were known: (i) the use of geometric relationships to select colours, e.g. *complementary colours*; (ii) *balancing the strengths* of colours, as indicated by the quote of Le Conti.

While Goethe's theories received very mixed support in scientific circles, anecdotal evidence and discussion of his work suggest that his theories do have some validity. A system based on Goethe's ideas was popularised during the 1980s for selecting clothing colours based on the wearer's skin tones (Jackson, 1985). Despite significant changes in fashion since its initial publication, it is still very popular, being placed in the top fifty most popular books in Amazon.com's "Beauty and Fashion/Cosmetics" category when sampled between May and June 2009. This would imply that the colour relationships embodied in the system transcend fashion, a supposition supported experimentally, with St.Bernard (1995) reporting that females whose clothing colours harmonised with their personal colouring are viewed more positively.

Goethe's writing on the emotional aspects of colour is also the basis of a psychological profiling system based on the subject's colour preference (Lüscher, 1970) and while the soundness of the approach may be questioned, the association of emotion with colour is unquestionable: as Gage states *"Nevertheless, the Lüscher system certainly rests on what seems to be a universal urge to attribute affective characters to colours, and it must be taken at the very least to be a modern manifestation of that urge" – Gage (2000).*

"Newton characterised light; Goethe contemplated appearance" state Jacobson and Bender (1996) and in the design of colour schemes, appearance is the dominant concern. It would therefore seem appropriate to use colour spaces based on perception when deriving harmonious colour schemes rather than those based on physics. As will be seen, this does turn out to be the case.

[21] Refrangible – capable of being refracted.

2.8 Perceptually-based colour arrangements

Goethe was interested in the emotional effects of colours and their combinations. These are the same effects that are important when creating an interface colour scheme. In Goethe's chromatic circle, the colours are placed around the perimeter in accordance with their perceptual complements, and while Goethe acknowledged black, white, and the greys as being separate from the colours, he did not attempt to align each colour with others of similar lightness. His arrangement is perceptually-based only with regard to hue; variations in saturation are not represented.

Two arrangements that have lightness as a third dimension have had popular support. The dual cone and the sphere both have lightness as the vertical axis, and the hues arranged around the equator. An example of the spherical arrangement by Runge is shown in figure 2.16. It combines the properties of a circular arrangement of hues, like the colour wheel, with the third dimension of lightness, with colours of varying saturation being found radially between the equator and the central vertical axis.

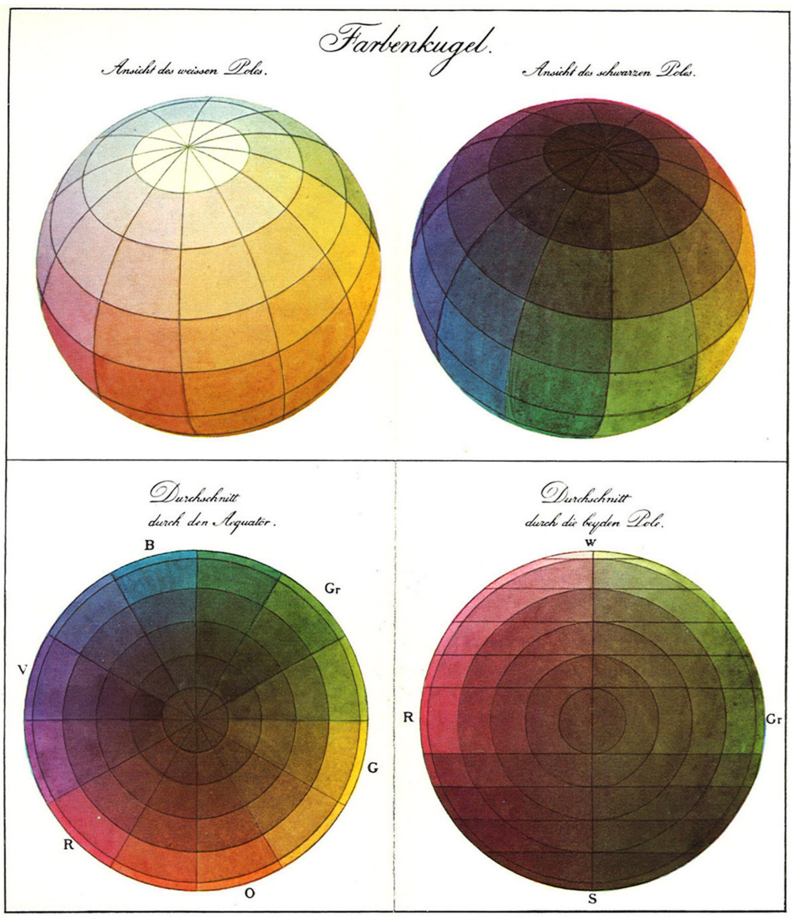


Figure 2.16: The arrangement of colours as a sphere by Runge (1810). The spherical arrangement is an idealisation, and is not derived from measurements of perception.

2.8.1 Munsell's colour order system

Albert Munsell, an artist and a teacher, having been frustrated by the use of meaningless colour names (e.g. "invisible green", "widow's joy", "elephant's breath", and "rat color"), was interested in methods of organising and clearly communicating colours. He used published spherical arrangements of colours in his early work and clearly acknowledges Runge's sphere, but notes that Runge's placement of red–green–blue at 120° spacing does not agree with perceptual measurements (Munsell, 1905). In Munsell's colour sphere of 1900, he arranged the colours with the pure hues positioned at the same height as a grey of the same lightness, which he called *value*. This is a generalisation of Goethe's colour wheel as it now includes lightness, but unlike Runge's sphere, it is based on perception.

In later experiments, Munsell used Maxwell's spinning disks with differing proportions of each colour to measure the relative lightness and placement of colours forming complements. The result was the first complete perceptually-derived arrangement of colours. The vertical position of each colour is determined by aligning it with a grey of equivalent lightness, and the radial distance corresponds to colour strength. Munsell's Colour Order System, and the description of its derivation in the paper "A Color Notation", was the first arrangement of colour derived from how humans *perceive* colour. It was seminal work in the perceptual measurement of colour (Munsell, 1905, 1907, 1912; Munsell and Farnum, 1941).

Munsell came to realise that a colour sphere, while ideal for illustration, does not correspond with the arrangement derived from measurements of colour perception:

> *"Desire to fit a chosen contour, such as the pyramid, cone, cylinder or cube, coupled with a lack of proper tests, has led to many distorted statements of color relations, and it becomes evident, when physical measurement of pigment values and chroma is studied, that no regular contour will serve."*
> *– Munsell (1912), p239.*

Munsell's Colour Order System has a lightness axis vertically, chroma (the strength) of the colour radially, and hues at various rotations. Taking lightness into account automatically precludes the shape from being spherical, as can easily be demonstrated by considering yellow. In its purest (most saturated) form, yellow is nearly as light as white. Therefore, full saturation yellow should not be on the same level vertically as pure red, green or blue, all of which are darker. Arranging the colours thus gave a most irregular shape which Munsell named a "color tree" (fig. 2.17). As this colour space was based on measurements from human perception, the units of the space reflected the perceived differences, not the idealised ones.

The colour space resulting from Munsell's experiments is known as the *Munsell Colour System*. The specifications of the colours are distributed as small colour samples in the *Munsell Book of Color*, printed to ensure the colours are accurate. While the colour values are now known in CIE coordinates and as spectral data (source cited in Romney and Indow (2002)), products based on Munsell's Book of Color are still

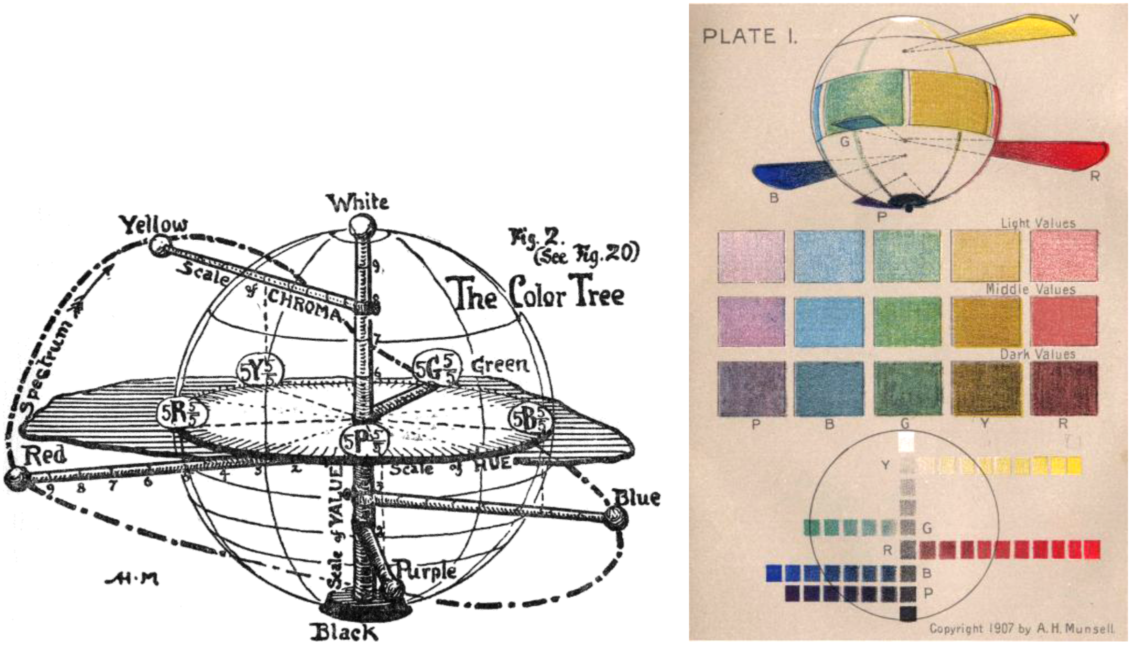


Figure 2.17: Munsell's illustrations from "A Color Notation" (Munsell, 1907) showing the placement of colours within his Color Tree. Left: an illustration showing the colours arranged taking into account lightness (vertically), and chroma (radially), and with the hues positioned to reflect the complementary relationships. Right: an illustration showing that, compared to red and green, pure yellow is lighter, while blue is darker. At the shown rotation, red has the maximum chroma (colour purity or intensity).

available commercially[22].

The scales of the Munsell colour space are numeric. The value axis (corresponding to lightness) is measured in unit steps from zero to ten, chroma (saturation) also in unit steps, and hue specified using one of five hues: red, purple, blue, green, yellow or mixtures of adjacent pairs as shown in figure 2.18. Colours in the Munsell system have a standardised syntax: <hue><value>/<chroma>. The hue is a gradation from 1–10 of the dominant hue(s) followed by the value and chroma, such as "5R 7/4" – a red with medium high value (7) and quite desaturated (4). The maximum chroma depends on both the hue and the lightness of the colour being measured.

2.8.2 Measuring colour difference and perceptual uniformity

It is useful to be able to measure colour differences when designing interface colour schemes. Colour difference could be used (for example) to ensure readability or distinguishable colour areas by making the distance in colour space between two object sufficiently large. Ideally, the measure would not depend on the colours being compared. A colour space in which the apparent change in colour is the same when moving from a point p to $p + \delta x$ for any initial point p and movement δx is known as a *perceptually uniform* colour space. The most well-known approximation to a perceptually uniform

[22] `http://www.xrite.com/top_munsell.aspx`, accessed May 20, 2010.

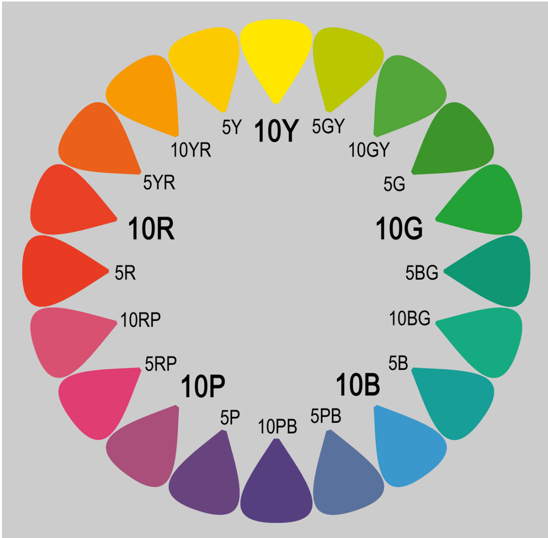


Figure 2.18: A colour wheel derived from Munsell's placement of hues, illustrating the five main colours, their interpolation, and their naming. Each of five primaries (red, yellow, green, blue and purple, can be subdivided to give colours between the primaries. The naming is from 1–10 of each primary or mixed pair of primaries, e.g. 10Y, 1GY, 2GY – 5GY – 10GY, 1G.

colour space is Munsell's Colour System Birren (1969a). Later studies have brought to light regions of non-linearity (Kuehni, 2001, 2003), but Munsell's colour space is a close approximation and is widely referred to as being "perceptually uniform".

Munsell's work resulted in a numerically-specified colour space based on human colour perception, not on the physical properties of light. Moreover, while colours are specified as numeric coordinates within Munsell's colour space, there is no simple mathematical transform to convert these coordinates to points in any of the CIE or RGB colour spaces; the conversion requires lookup tables.

To provide a colour space in which Euclidian distance is more closely aligned with perceptual difference and is mathematically tractable, the CIE has defined some colour spaces that, on one hand, are approximately perceptually uniform and, on the other, can be transformed into CIEXYZ coordinates (and therefore into all the other colour spaces for which transforms from CIEXYZ exist). In order to understand the basis upon which these spaces were constructed, it is necessary to understand a little of the human sensitivity to colour difference.

2.8.3 Human colour sensitivity and discrimination

In good lighting conditions, humans are capable of discerning a large number of distinct colours. Judd (1959) estimated the number of discriminable colours at ten million. More recently, McCamy (1998) gives the number of discernable colours as 1,875,000 under normal conditions, increasing to 7.5 million under good viewing conditions.

Within the range of visible colours, the sensitivity of human colour vision to changes in wavelength is not uniform. Two items the same distance apart in terms of wavelength may be distinguishable in one part of the colour space, but appear indistinguishable in another. The *just-noticeable-difference (JND)* is the minimal difference between two perceptual quantities for there to be a *perceived* difference. MacAdam (1942) measured the chromaticity differences necessary for distinguishability and found them to be unevenly distributed: the perceived differences depended on the colours being compared. Within a perceptually uniform colour space, the set of colours that differ from a given colour by less than the JND would occupy a sphere. As MacAdam was working with the CIEXYZ colour space, which is perceptually non-uniform and is a 2-dimensional projection, the spheres map to ellipses, and JNDs are therefore often referred to as MacAdam's ellipses. Later studies, such as that of Noorlander and Koenderink (1983), have confirmed the validity of the results, and Alleysson and Hérault (1997) has shown that the effect can be largely attributed to the non-linearity between the inputs and outputs of the photoreceptors in the eye.

As with many aspects of colour perception, the data is somewhat subjective: there are individuals for whom the ellipses are not only the wrong size, but are also oriented (compared to the norm) in the wrong direction. It is well-known that colour preference is subjective. The reason for some of the subjectivity becomes clear when the scale of individual differences, which are not often evident in the aggregated results, are taken into account. MacEvoy (2009b) discusses this at length and quotes the following two sources:

on the uniformity of MacAdam's ellipses:

> *"When one confines the intercomparison [between subjects] to a particular location in the chromaticity diagram, the ellipses of different observers are quite often not in close agreement. Rather, larger discrepancies are noted in the orientation, size and shape of the ellipses." – Wyszecki and Stiles (1982, p323) as cited by MacEvoy (2009b).*

on hue perception:

> *"Comparisons between different observers, whether in the same or a different experiment, present a discouraging picture. Although observers agree on certain major trends, individual differences are best described as enormous." – Kaiser and Boynton (1996) as cited by MacEvoy (2009b).*

Kuehni (2003) examines the topic of small colour differences in depth, and in a later work (Kuehni, 2004), reports on the large variability in *unique hues* (hue without any hint of their neighbouring opponent colours, such as a pure blue, without any tinge of red or green). Figure 2.19 illustrates the range of colours perceived as being the primary "pure colours", with green occupying nearly one quarter of the range.

Fortunately, in spite of these very large individual differences, there is sufficient overall agreement about the identities of the perceptual primaries, and measurements of colour difference are sufficiently constrained, to ensure that the idea of a typical

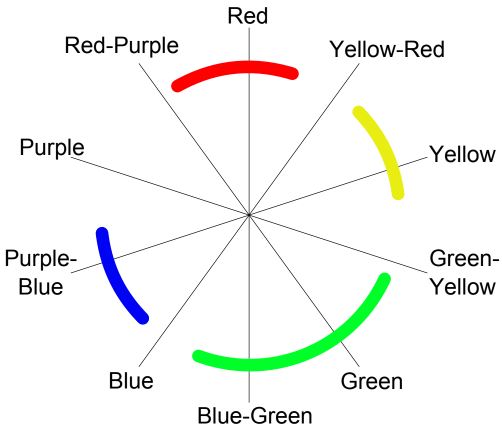


Figure 2.19: An indication of the variability in human colour perception. Overlaid on the Munsell hue circle are coloured bands indicating the range of colours perceived as the pure primary colour. Derived from Kuehni (2004).

human observer is meaningful. However, the variability could explain (in part) why colour preference is subjective.

2.8.4 The CIELAB colour space

To standardise the measure of colour difference, various transformations of the CIEXYZ colour space have been proposed (as mentioned in sec. 2.8.2). The most widely used is the 1976 CIELAB colour space Fairchild (1998) defined by the CIE. It has coordinates $L*$ (for lightness) ranging from 0 to 100 (black to white), and chromaticity axes $a*$ and $b*$. The transform from CIEXYZ to CIELAB is given by:

$$L* = 116\left(\frac{Y}{Y_w}\right)^{\frac{1}{3}} - 16 \quad (2.5)$$

$$a* = 500\left[\left(\frac{X}{X_w}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_w}\right)^{\frac{1}{3}}\right] \quad (2.6)$$

$$b* = 200\left[\left(\frac{Y}{Y_w}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_w}\right)^{\frac{1}{3}}\right] \quad (2.7)$$

(2.8)

where:

X, Y, Z – the CIEXYZ value of the colour

X_w, Y_w, Z_w – the CIEXYZ value of the colour being used as the white-point.

In the case of extremely dark colours, an alternative set of equations is used, see (Fairchild, 1998). The asterisks are commonly dropped when there is no chance of ambiguity. For the derivation of the model, including details of the multiple transform steps embodied in the overall transform, see MacEvoy (2009c).

The CIELAB colour space is an absolute colour space and has defined mappings to and from CIEXYZ, but not RGB. As previously explained, any RGB colour space is device-dependent: the actual colour depends on the device, so it not possible to specify transformations to or from an RGB space without knowing the characteristics of the device with which it is associated. Once these characteristics have been defined (as they have been for sRGB, for example), then colour values can be transformed from that RGB space via CIEXYZ to CIELAB.

The CIELAB space has chromaticity axes and a lightness axis. Therefore, there is an implicit "white" – the colour that will be seen when $L = 100$ and $a = b = 0$. As there is no perceptually absolute "white", it is necessary to choose one, so the CIELAB to CIEXYZ transforms require the specification of a white-point (usually 5000K or 6500K). When transforming to sRGB, the 6500K white point should be used. If this is not done and the white point left at 5000K, the transformed value of Lab = (100, 0, 0) (white) will have a yellowish tint when displayed on an sRGB display. This effect disappears using the 6500K white point.

ΔE – a measure of colour difference

A measure of colour difference can be useful to determine if two colours are likely to be indistinguishable (have negligible colour difference) or, as could be useful in colour scheme design, to ensure that two items have a sufficiently large colour difference to appear differently coloured. RGB colour spaces are significantly perceptually non-uniform, so there is no consistency in the perceptual difference between a pair of colours separated by a Euclidean distance δx in one region of the RGB colour space and another pair of colours δx apart in another region of the colour space. It is more useful to measure colour differences in a perceptually uniform space, such as CIELAB. The *perceived* differences would then be linearly related to the distance in colour space.

A common measure of colour difference is ΔE, defined by the CIE in 1976, which measures distance in the CIELAB space. There have been later colour difference formulae that have attempted to correct for non-linearities in the CIELAB space or are tailored for specific application areas, but the 1976 ΔE is still the most widely used. For two colours c_1 and c_2:

$$\Delta E = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2} \tag{2.9}$$

$$\Delta L = |L_1 - L_2| \tag{2.10}$$

where:

ΔE – the colour difference – the Euclidean distance is CIELAB space

ΔL – the lightness difference, which is important to the readability of text.

L_i, a_i, b_i – for i = 1, 2 – CIELAB coordinates of colours c_1 and c_2

Both of these measures of visual difference are useful in the context of the user interface colour scheme design. ΔE, the overall colour difference, could be used to ensure easily distinguishable interface objects, and ΔL, the difference in lightness, is important for ensuring readability, the requirements for which are detailed in section 2.9.8.

Ostwald's colour space

Ostwald theorised that *order leads to harmony* and derived a colour space based on opponent colours and the pre-existing idea that any colour could be visualised as the pure hue mixed with proportions of black and white. His work (Ostwald, 1969), was later adopted as the basis of the Swedish NCS system (Scandinavian Colour Institute, 2009):

> *"NCS – Natural Color System®© is a logical colour notation system which builds on how the human being sees colour. A notation represents a specific colour percept and describes the colour visually, it is not depending* [sic] *on limitations caused by pigments, light rays or nerve signals that have given rise to this perception."*[23]

The specification of colour using NCS is easier to explain than spectral, RGB, or three dimensional hue/saturation/lightness representations. It has two pairs of opponent hues (red/green and blue/yellow) and a vertical lightness axis in a double cone arrangement. Forty hues are defined (fig. 2.20[24]), arranged around the centre of the two cones. The internal colours are defined as mixtures by adding varying ratios of black and white (fig. 2.21), a property that has an attractive conceptual simplicity. The NCS system is perceptually-based, but not perceptually uniform (Kuehni, 2003). The NCS is widely used and while it is possible to derive harmonious colour combinations from paths in the colour space (Hård and Sivik, 2001), its definition in terms of colour samples and its lack of perceptually uniformity (Kuehni, 2003) limits its usefulness in the automatic design of colour schemes.

2.8.5 RGB colour spaces

For reasons of computational efficiency, it is desirable for the native colour values in a computing system (and on the Internet) to be usable without significant transformation by the display device. If, for example, CIELAB were used as the native colour representation, given that displays use RGB values, every pixel sent to the display would need to be transformed from CIELAB to RGB. As the transform uses power functions, for this to happen sufficient quickly, all machines, from the cheapest netbook up, would need to include graphics co-processors, with the associated increase in cost and power consumption. Similar problems would occur using NCS or Munsell spaces, with the need for lookup tables and interpolation.

[23] From the "A Visual Colour System"page accessed via the "The NCS System" menu item on `http://www.ncscolour.com` (accessed April 16, 2009).

[24] NCS – Natural Color System®© property of Scandinavian Colour Institute AB, Stockholm 2009. References to NCS®© in this publication are used with permission from the Scandinavian Colour Institute AB. NCS images from `http://www.ncscolor.com` – reproduced with permission.

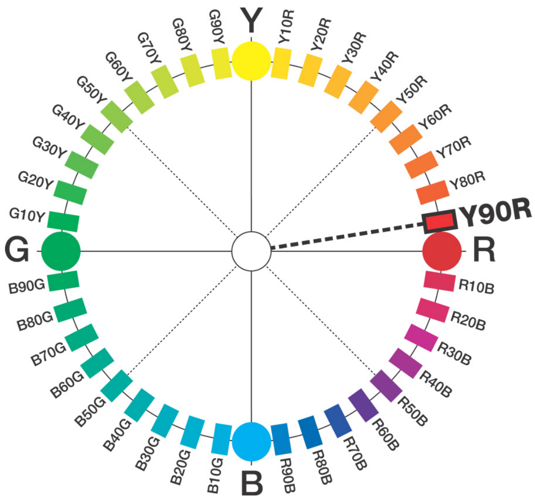


Figure 2.20: The NCS colour system uses forty experimentally derived hues, with the opponent colours having prominent places.

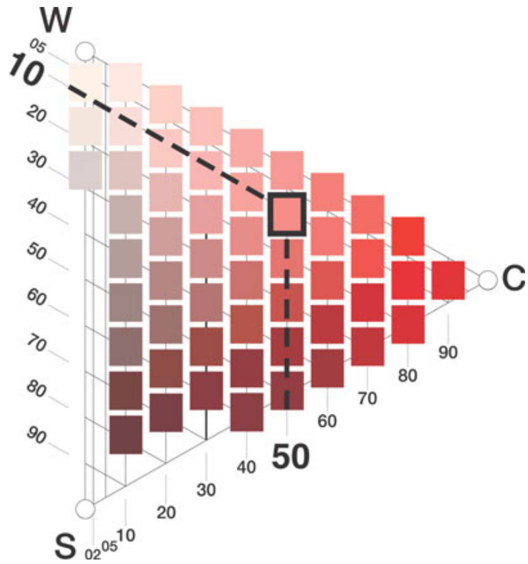


Figure 2.21: A vertical slice of the NCS colour system gives the colours of a single hue (Y90R) mixed with black and white. The full saturation colour is always at the same vertical position, irrespective of its lightness.

While an absolute colour space such as CIELAB would provide simple colour management and correspond well with human perception, the increase in cost is not justified. Instead, given that RGB tristimulus values are sufficient to represent a colour, it comes as no surprise that RGB values are used as the native colour representation in all common operating systems, instead of an absolute representation of colour. This benefit comes at a price. While the idea of specifying colours using three values does require some familiarisation, the relationship between RGB values and the perceived colour is abstruse and therefore quite unsuitable for use by end-users. The raw RGB values do not transform easily to the more comprehensible dimensions of hue–saturation–lightness, as used by the Munsell or NCS colour spaces, or the CIELCH (Lightness, Chroma, Hue) colour space, a polar form of the CIELAB space.

To enable RGB colour values to be manipulated in a more meaningful way without losing the speed advantage, there are two polar transformations, HSB (hue, saturation, brightness), and HSL (hue, saturation, lightness). There is, however, a problem with using HSL and HSB for colour scheme design: the lightness of colours varies widely as the hue is changed. These colour spaces may be convenient mathematically and conceptually, but not being perceptually-based, they are not ideal colour spaces for colour scheme design. Nevertheless, as they are easy to understand, both are widely used, with the Microsoft *Windows*™ colour selection dialog having both RGB and HSL modes.

2.9 Abstract representations of colour harmony

In order to understand colour harmony, it is useful to trace the evolution of the concept and to explore some theories about how colour harmony can be produced. These ideas could be addressed from two perspectives, either the neurophysiological approach suggested by Martindale (2001) who proposes a neural model of aesthetics, or the pragmatic guidelines that have been found to result in pleasing colour aesthetics.

The guidelines that have been refined by the artistic community are widely used, and are promoted as being helpful in deriving pleasing colour schemes. What is of interest in this research is the use of computation to derive harmonious colour schemes, a much more modest goal than computational aesthetics.

The artistic guidelines and colour models based on perceptual phenomena are widely used to teach the principles of colour harmony, are well understood and have been refined over many years by the artistic community. They have a long history of successful use in the derivation of harmonious colour schemes. It is therefore appropriate to use these models as the basis for the derivation and evaluation of colour aesthetics in this research, rather than a more general and less directly applicable neurophysiological model.

In order to understand how to derive harmonious colour schemes, it is necessary to define what is meant by *colour harmony*, what specifically is meant by a colour scheme, and to consider how such schemes may be derived algorithmically.

2.9.1 The concept of colour harmony

Colour harmony is an intuitively simple idea that might be phrased as *a set of colours that look good when seen together.* To determine its dimensions, more precision is needed, but obtaining a consensus on a more precise meaning is difficult.

Various author's contributions are:

- Sully (1879), writing on colour harmony between the time of Newton and Munsell, clearly shows both the diversity of thought at that time (such as attempts to analogise colour and musical harmony) and the difficulty caused by the lack of any precise means of colour specification. The paper is 19 pages long, but does not attempt to precisely define colour harmony. The closest approaches to definitions are the following: *"accounting for the agreeable and disagreeable effects of combined colours"*; *"some of the most delicious effects of colours in combination"* and *"that complementary colours have a special aesthetic value seems indisputable"*;

- *"the 'suitability' of juxtaposed colors" – Polzella and Montgomery (1993)*;

- *"selection of colors that give pleasure"* and *"grouping of colors to suit some practical use" – Munsell (1905)*;

- *"colors seen together to produce a pleasing affective response are said to be in harmony" – Burchett (2002)*;

- *"in the theory of music, analogous relationships between tones have been studied and formulated for centuries. By contrast, in the visual arts the rules for color relations have been limited to a few suggestions on how to obtain 'harmony', that is, colors that do not clash." – Arnheim (1987)*;
- *"when two or more colours seen in neighbouring areas produce a pleasing effect, they are said to produce a colour harmony" – Judd and Wyszecki (1963) as cited by Ou (2006).*

"Scores of books giving the opinions of experts have been written on color harmony. Contradictions in these opinions are frequent" write Judd and Wyszecki (1963) (cited in Rapoport and Rapoport (1984a)). The scope of the problem is well summarised by Burchett (2002) who states:

> *"Color harmony is learned, which has been known for many years; yet despite important evidence for this, no acceptable model exists for explaining the concept of color harmony. When two or more colors are brought together to produce a satisfying affective response, they are said to be harmonized. But very little is known about why such an effect occurs. Color harmony has a wide range of meanings. It is an abstruse concept described differently by different authorities and has diverse meanings when applied to the process of making art and when used in art-related and other specialized color-use opportunities."*

2.9.2 Necessary conditions for colour harmony

In an attempt to clarify the dimensions of colour harmony, Burchett (2002) performed a linguistic analysis of the usage of various terms associated with colour harmony in twelve books on the use of colour. The findings were not clear-cut, but did result in useful information:

> *"The concept of color harmony exists with different meanings. The study failed to demonstrate a general acceptance of any ranked list of terms. . . . Yet the predominant understanding of color harmony was determined to be attributed most frequently to order, referring to uniformly spaced points in a color classification system." – Burchett (2002)*

This is consistent with other authors, who have found order to be important. Munsell (1905) says: *"Let us leave these musical analogies, retaining only the clue that 'a measured and orderly relation underlies the idea of harmony' "*; and Ostwald (1969) uses no principle other than order in the derivation of harmonious sequences.

Moon and Spencer, in their proposal for a model of colour harmony, put this quite succinctly: *"the basic principle behind all this work is that any arrangement of colors that can be sensed as an orderly combination will be pleasing"* – Moon and Spencer (1944c). While their overall model (Moon and Spencer, 1944a,b) was later found to have poor predictive performance (Granger, 1955a), they too highlight the necessity for order and also for unambiguous colouring[25]: *"it is a fundamental principle of aesthetics that the observer should not be confused by the stimuli: there should never be a feeling of uncertainty" – Moon and Spencer (1944c).* This is consistent with the use of complementary colours which, by definition, must appear different, although surprisingly the colours with the greatest contrast are not always complements (Mahyar et al., 2007) .

As noted in Ou (2007) and Westland et al. (2007), two recent comprehensive reviews of colour harmony literature, two principles appear to dominate discussions of harmonious colour selection. These are order and, as already noted (section 2.7), the use of complementary colours. Before considering how harmonious colours might be selected, it is necessary to consider more closely what is meant by "order".

Visually apparent order

As previous stated, order is a dominant concept in colour harmony. This author would strengthen *order* to *visually apparent order*. The distinction is subtle. If there is a visual progression in the placement of objects, for example, they are perceived to form a row, a column, or any perceptible line, it is important that the sequence of colours derived from a progression in the colour space proceed in step with this perceived line. The need for this was noted during some experimental work in which a sequence of colours was used to colour a vertical set of buttons on a user interface. When the buttons were coloured in sequence with ordering of colours from the colour space (e.g. along a complementary line, or a linear progression in lightness) the effect was harmonious. When the colouring had discontinuities or contained multiple sequences (e.g. a split-complementary scheme), the adjacency of apparently unrelated colours was not harmonious. However, the same colours, when not part of a visually apparent progress, do harmonise.

2.9.3 Harmonious colour selection using colour wheels

Newton's and Goethe's colour wheels are among many that have been used to select complementary colours. That the diametrically opposing colours are not necessarily perceptually complementary (that is, they do not necessarily blend to give an achromatic colour) appears to make little difference. As noted in MacEvoy (2009a), the term "complementary" has several interpretations, with Goethe's after-image complements not quite aligning with Munsell's additive complements. Evidently, great precision is not required when choosing the colours to be used as "complements" within a colour scheme. Another interpretation could be that it is less necessary for the colours to be complementary than to have unambiguous chromatic contrast. Selecting approximately diametrically opposed colours from any common colour wheel will produce a pair of colours that have sufficient chromatic contrast to be used as "complementary" colours.

It has been found that harmonious colour schemes can often result from the colours at the extreme points of geometric shapes placed on a colour wheel, as illustrated by the schemes (excluding the monochromatic scheme, which will be considered shortly)

[25] unambiguous colouring – colours that are either sufficiently similar that they are obviously related, or sufficiently distinct that they are obviously *not* related.

in figure 2.22. Colours that have a small angular separation on the colour wheel can also appear harmonious, as there will be a visual similarity to the colours. A scheme that conforms to this requirement is termed “analogous”.

Rotating the shape will cause the colours underlying the indicated points to change, giving differing sets of colours, all broadly conforming[26] to the colour relationships defined by the shape. However, as the colours selected using such a method are from the end points of the shapes (see fig. 2.22 omitting the monochromatic scheme), even allowing for rotation, the colours will only vary in hue, and so will only cover a very small subset of the colours available, because of the lack of variation in saturation or lightness. Unfortunately, this is consistent with much of the discussion on harmonious colour schemes in the popular literature: hue dominates the discussion, with little reference being made to either lightness or saturation, a situation that was noted by Westland et al. (2007).

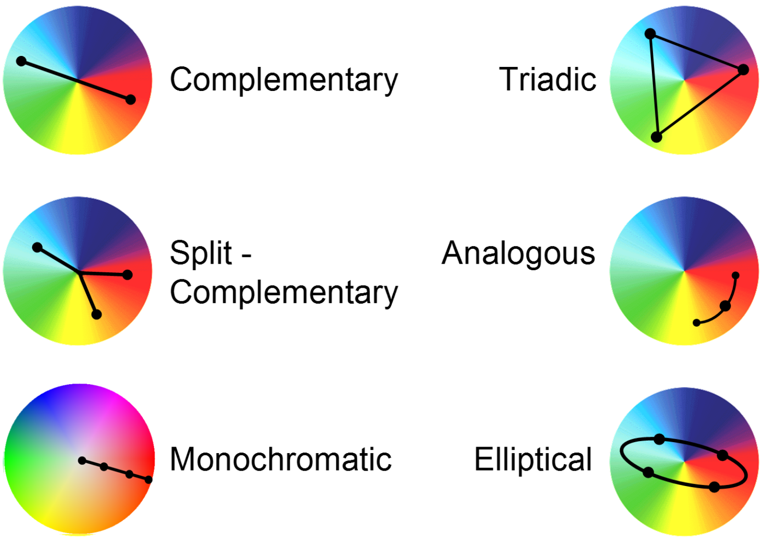


Figure 2.22: Examples of different types of colour schemes that can be derived using geometric shapes and a colour wheel. The monochromatic scheme is not viable unless, as shown, the colour wheel is modified to vary lightness or saturation (or both) along the radial axis.

To enable a wider range of colours to be selected using a colour wheel, it is necessary to include variations of either lightness or saturation. A widely used method is to place one of the colours white, grey or black at the centre of the disk, and interpolate between the centre and the fully saturated colours at the edge of the disk. The colours would then vary in the radial direction with the lightness or saturation, or a combination of the two, as a function of the radial distance. Such an arrangement could be named a *colour disk* and, while it allows for a much wider range of colours than the full saturation colour wheel, the colours are still a small subset of those available. Using a colour disk, with a radial variation of lightness or saturation or both, the monochromatic scheme becomes viable (fig.2.22) and it can be used to create simple and elegant colour schemes.

[26] the actual colours will depend on how the wheel is constructed.

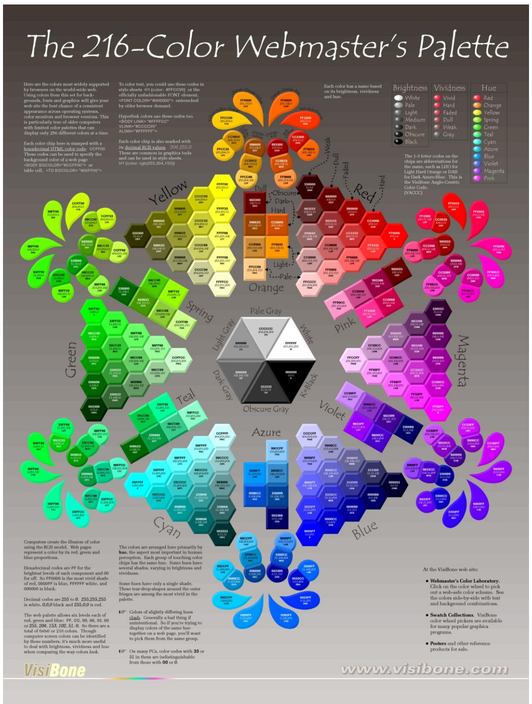


Figure 2.23: An circular arrangement of colours modelled on a colour wheel that shows variations of saturation, lightness and hue, from *http://www.visibone.com/color/poster2x.html*, accessed May 21, 2010.

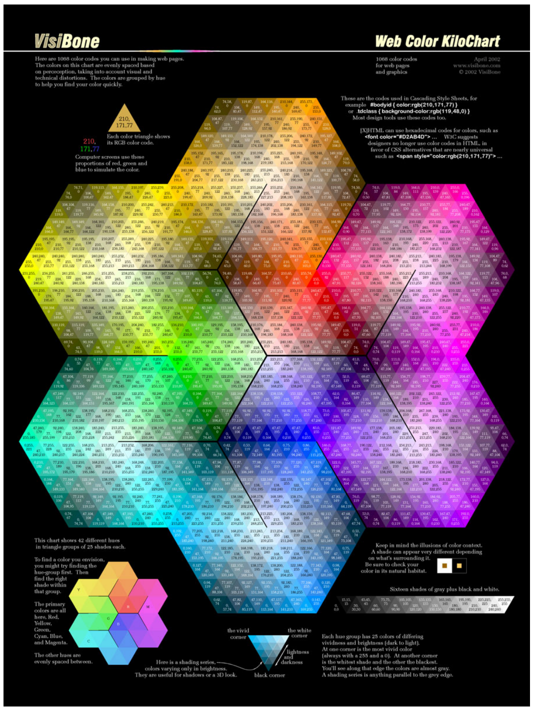


Figure 2.24: An alternate expanded 2D visualisation featuring perceptually uniform spacing, from *http://www.visibone.com/color/kilochart%5F900.jpg*, accessed May 21, 2010.

This form of colour wheel is widely used to derive sets of harmonious colours (Ball, 2001; Lyons and Moretti, 2004). There are variations that display colours from all three colour space dimensions to aid in colour scheme design, such as the innovative Visibone diagrams shown in figures 2.23 and 2.24[27], but to ensure that all colours are available, a three dimensional model is necessary. Itten, a long-standing teacher of art, describes many geometric shapes based on two and three-dimensional shapes in colour space, extending the set shown in figure 2.22 (Itten, 1970a).

2.9.4 Interpolation in colour space

With the addition of either lightness or saturation as a radial dimension, the colour disk can be used for selecting many more colours than are possible when selection is limited to the fully saturated colours. By interpolating between any two points on the shape, subdividing the line, and using the points to indicate a colour, an ordered progression of an arbitrary number of colours may be selected. Combining the shape vertices with interpolated positions provides a means of choosing an arbitrary set of

[27] Figures 2.23 and 2.24 used with the kind permission of Bob Stein of `http://www.visibone.com`. For more details on figure 2.23, see `http://www.visibone.com/color/poster.html` and on figure 2.24, see `http://www.visibone.com/color/kilochart.html`. All URLs accessed May 21, 2010

colours that conform to the key colours from shapes found to yield harmonious colours, and provide the essential requirement of a visually apparent order to the colours.

The position of the hues around the circumference is not fixed. Many variations can be used: RGB (red–green–blue) or RBY (red, blue yellow) at 120° spacing; variations on the opponent colours (NCS); the Munsell colour space; or any of a very large number of other arrangements, depending on the intended application. Different colour wheels arrangements are needed for diametrically opposing colours to be blended achromatically, depending on whether the mixing is additive (light is added, as with displays), or subtractive (as with paints and dyes, where the colour absorbs light). The subject of colour positioning on a colour wheel is discussed in great detail by MacEvoy 2009a; 2009b who has the great advantage of having both a scientific and artistic understanding. His extensive website[28] clarifies many points of confusion between the artistic and the scientific use of color.

Extending the colour wheel to three dimensions by adding a vertical lightness axis with the full saturation hues being aligned with mid-grey, and saturation varying radially, gives the arrangement known as the HSL (hue-saturation-lightness) colour space (fig. 2.25), or the HSB (hue–saturation-brightness) space, in which full saturation colours are aligned vertically with white (fig. 2.26). It is often difficult to visualise the inside of such colour solids. The cutaway view of the HSL space (fig. 2.27) inspired by a graphic in Varley (1980), helps in visualising the internal colours.

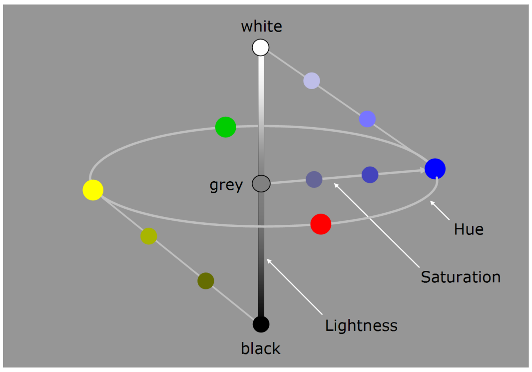


Figure 2.25: The HSL (hue-saturation-lightness) colour space is biconical with the full saturation colours aligned with mid-grey.

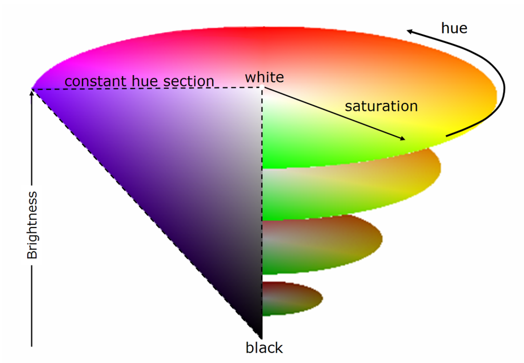


Figure 2.26: The HSB (hue–saturation-brightness) space aligns the full saturation hue with white.

Wireframes: the word "shape", when referring to a geometric shape representing a set colours to be selected from a colour wheel or disk, is rather generic, so the term "wireframe" will be used instead. This embodies the idea of rigidity, of an outline rather than a volume, and is easy to visualise. While helpful as a starting point (especially for non-artists), there is much that is *not* specified when using wireframes to select colour schemes. If the wireframes are restricted to colour wheels (as in common in the popular literature), then the selectable colours can only ever be a subset of those

[28] `http://www.handprint.com` (accessed May 21, 2010)

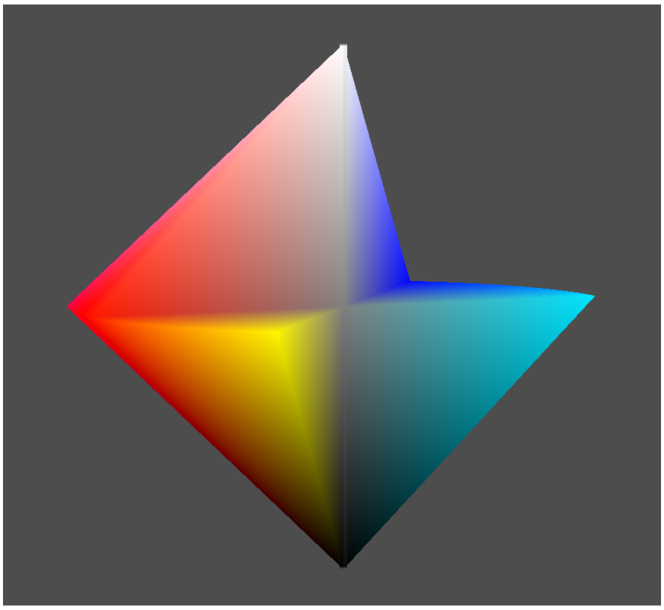


Figure 2.27: A 3D view of the hue-saturation-lightness colour space, with the cutaway section allowing a view into the interior of the space (from software by the author, inspired by a diagram in Varley (1980)).

possible. The colours on a colour wheel are frequently fully saturated colours. Such colours are rarely used by professional designers. Even with the wireframe positioned in the 3D HSL colour space, so that colours of varying lightness and saturation are possible, the simple heuristic of selecting colours using a wireframe on a colour wheel can, but does not always, yield harmonious colour schemes.

Colouring a website or a graphical user interface using this method will be found to produce very mixed results. This appears to be at odds with the expectations of the successful artists who have promoted the colour wheel/wireframe method as a good way to introduce colour harmony (Graves, 1952; Itten, 1970a; Howlett, 1996; Wills, 1997; Coulson, 1998; Foster, 2004). There must be a factor missing when applying the technique, or the heuristic is being misinterpreted.

2.9.5 The difference between a colour scheme and a palette

The idea that some sets of colours work well together is widely accepted, even if the means of selecting them is not. Sets of colours – palettes – are regularly featured in magazines about fashion and interior design. To help designers with colour selection, a myriad of publications exist, from the pamphlets and books published by paint manufacturers to books of harmonious colour schemes (Whelan, 1997; Sawahata, 1999; Krause, 2002; Walch and Hope, 1995; Barker, 1999) and more recently web sites devoted to colour palettes (Adobe Systems, 2009; Monsef IV, 2009; Morton, 2009b).

It might appear that the selection of a good colour scheme involves nothing more than the selection and application of a palette by a skilled designer to the objects to be coloured. Unfortunately, this does not always lead to a pleasing result. The reason for this lies in the difference between a palette and a colour scheme.

A palette is a predefined set of colours, whereas a colour scheme is both a set of colours (a palette) and a binding of those colours with objects to be coloured: *"pattern of colors or colored objects conceived of as forming an integrated whole"*.[29] This association of colours with specific objects allows the use of information about objects, such as their spatial layout and size, to influence the choice of colours in the palette. The colours in a palette will have been selected to colour a set of objects whose size and spatial relationship are known. If a palette is used to colour objects whose size and arrangement are significantly different from those envisaged by the palette designer, the resulting effect is unlikely to be as pleasing as the original.

The relationship between an object's size and its colour

When one examines the work of artists (e.g. Albers, 1963; Hornung, 2005) or graphic designers (e.g. Wright, 1995; Howlett, 1996) more critically, it is evident that variations of lightness and saturations of the colours picked from a colour wheel are used.

Fully saturated colours, especially more than one, are usually only used to attract attention. In any application where a scheme will be viewed for extended periods, the colours will have been muted by tinting (adding white), shading (adding black) or toning (adding grey). The skill of a graphic designer is in understanding which hues should be used and exactly how those should be modified by altering their lightness and saturation for the best visual effect in the context of the other objects in the scene.

In the case of a colour palette shown in an interior design magazine, the palette is generally accompanied by a photograph indicating which colours from the palette are to be used for the ceiling, for each wall, and for other objects in the room. The designer will have considered the size of objects when choosing the palette colours and altering these associations is unlikely to provide as pleasing a result. It is therefore evident that the area and placement of colours influences the aesthetic effect.

2.9.6 Munsell's theory of colour harmony

While Munsell is best known for his work on his colour order system, he also published work on the selection of sets of harmonious colours. Central to his ideas of colour harmony was the idea of balance:

> *"What is balance of color? Artists criticise the color schemes of paintings as being 'too light or too dark' (unbalanced in value), 'too weak or too strong' (unbalanced in chroma), and 'too hot or too cold' (unbalanced in hue), showing that this is a fundamental idea underlying all color arrangements"*

> *"Let us assume that the centre of the sphere is the natural balancing point*

[29] Excerpted from the definition of "color scheme". *Dictionary.com Unabridged (v 1.1).* Random House, Inc. `http://dictionary.reference.com/browse/colorscheme` (accessed December 5, 2008).

> *for all colors ... then color points equally removed from the centre must balance one another. Thus white balances black. Lighter red balances darker blue-green. Middle red balances middle blue-green. In short, every straight line through this centre indicates opposite qualities that balance one another. The color points so found are said to be 'complementary', for each supplies what is needed to complement or balance the other in hue, value, and chroma." – Munsell (1905)*

Munsell then continues *"how lighter colors may balance darker colors, how small areas of stronger chroma may be balanced by larger masses of weaker chroma"*. His theory is that, for a balanced (in his view, harmonious) composition, the strength of each colour, as given by the product of chroma and lightness, should be balanced against area. This is known as the "Munsell law of colour harmony":

$$a_1(v_1 c_1) = a_2(v_2 c_2) \tag{2.11}$$

where:

a_i – area of object$_i$
v_i – value (lightness) of object$_i$
c_i – chroma (saturation) of object$_i$

The aim is to balance the colour scheme around mid-grey, which is the most natural balance point but, Munsell stressed, not the only one. It is, however, the most widely used.

It is widely accepted that balance is important; it was mentioned by Le Blon (Gage, 2000, p139 and sec. 2.7) at the time of Newton and echoed by later authors:

> *"balance, or a feeling of rest, is the first essential for good colour arrangements. It is the principle underlying the well-known 'Law of Areas' which states: Large areas of colour should be quiet in effect, while small amounts may show strong contrasts; the larger the amount used, the quieter the colour should be, and the smaller the amount, the more striking the contrast may become. These contrasts may be due to a decided difference in hue, in value, or in intensity" – Goldstein and Goldstein (1961, p197).*

It is notable that colours selected using this technique are frequently not the highly saturated colours often used by beginners, but the more subtle "no-name" colours:

> *"when first seen on the equator of the sphere this degree of red is likely to be rejected as untypical: yet it is the type which appears most frequently in beautiful combinations, while the extreme red generally taken as typical is usually absent, or if a small touch is introduced as an accent, it will be found to balance with a correspondingly large area of the weaker blue-green." – Munsell (1912).*

The simple balancing of area against saturation (without considering lightness) has become known as "law of areas" or the "inverse area rule" (Morriss et al., 1982). As will be seen from the experimental results below, this heuristic has strong experimental support.

Implicit in Munsell's theory is that the colours being balanced are selected from a path in his perceptually uniform space, these paths being the wireframe shapes already mentioned, elliptical paths and diminishing spirals (the chroma decreases with lowering value). The theory incorporates order, balance, and the size of objects, as well as the wireframe shapes found to be helpful in colour scheme design. Given a wireframe shape, this allows sets of possible pairs of colours to be determined.

Experimentally, Munsell's theory of colour harmony seems to stand up rather well (Granger, 1955b; Morriss et al., 1982; Rapoport and Rapoport, 1984b; Morriss and Dunlap, 1987), but it should be noted that most colour harmony evaluation experiments use only two or three coloured items, significantly fewer than would be typical in a graphical user interface.

Munsell's theory is attempting to balance what he called "colour strength", which is often interpreted as the perceptual "weight" of the colours. Munsell calculated the colour strength as the product of value and chroma. This is consistent with the weights of colours from Alexander and Shansky (1976), who found *"apparent weight is a decreasing function of value and an increasing function of chroma. These results support the earlier qualitative findings that 'dark' colors appear heavier than 'light' colors, while providing quantitative meaning to the terms dark and light."*. Also noted was the independence of hue: *"the findings are clear. Hue contributes little to the apparent weight of colors. Value and chroma are the primary determinants of color heaviness"*. This supports Munsell's expression of colour harmony in terms of value and chroma, without considering hue. This allows a colour scheme designed taking only area, chroma and lightness into account. The resulting scheme should be reasonably harmonious whatever colours (hues) are used.

In a trial balancing two complementary coloured areas, Morriss et al. (1982) found that the area–chroma balance strongly correlated with that theorised by Munsell, although the results, while still strong, were less uniform for those without an artistic training or background. They also found that young children seemed to prefer schemes that contradicted Munsell's guidelines. Their preferences were for what would generally be thought of as discordant schemes, a result agreeing with Palffy (1976). In a later experimental study oriented towards testing the effect of balancing lightness, Morriss and Dunlap (1987) found strong support for the chroma–area balance predicted by Munsell's theory, but not the value–area balance.

Linnett et al. (1991) concurred with previous studies in supporting Munsell's theory, but only when area was the variable. When chroma was the variable, subjects preferred the larger areas to have higher chroma – exactly the opposite of the predicted result, an effect that was *"attributed to simultaneous or successive color contrast effects"*. Rapoport and Rapoport (1984a) give experimental support for Munsell's rules for colour harmony, and note that females are more consistent in their colour preferences, and their selections have a higher correlation with those predicted by Munsell's law of colour harmony.

Locher et al. (2005) found that the concept of weight and balance was well understood even without artistic training: *"design-trained and untrained participants in experiment 1 were in good agreement as to the perceived weights of the large color area in different versions of five of the six compositions. This suggests that the color – area – weight relationship in the stimuli studied was salient to viewers regardless of their training in the visual arts"*. The concept of visual weight, while well understood, is not always synonymous with Munsell's product of value×chroma. Munsell's theory would balance a dark object and a light one of equal saturation, whatever the background colour, whereas the visual weight of each object will depend on the background colour. Against a light background, the dark object will appear to be heavier than a light one, as noted by Morriss and Dunlap (1987).

2.9.7 Other models of colour harmony

There have been several later models of colour harmony. Moon and Spencer's model (Moon and Spencer, 1944a,b,c), like Munsell's, balanced area against attributes of colours, but also incorporated regions of distinction to be avoided. This theory was found to be a less reliable predictor of colour harmony than Munsell's theory (Granger, 1953, 1955a), but is notable for introducing quantifiable regions of ambiguity into the model. In a recent study, Ou (2006) derived a model determining the contributions to colour harmony of each of chroma, lightness and hue. Several findings are useful: blue was found to be the easiest colour to harmonise and red the most difficult. Three effects related to lightness were noted: colours varying only in lightness led to harmony; lighter colours appeared more harmonious; and small differences in lightness ($\Delta E < 15$) impair harmony. Order did not necessarily lead to colour harmony, although this could be related to the fact that the trial used pairs of colours. Colour order is more apparent when three or more items are being assessed.

While colour harmony guidelines can give useful results, the results are also dependent on the composition. Rapoport (1994) reports on the results of an experiment in which colour harmony guidelines from Munsell and Itten broadly agreed with each other, but not the experimental data. This was thought to be due to the lack of allowance for the composition being coloured.

The trends in these research findings are not completely consistent with each other, but are all consistent with the notion that a harmonious colour scheme will incorporate an ordered set of colours with a balance between the areas of colour and their saturation, and possibly their lightness, although the need for the inclusion of lightness is less certain. The use of prototypical colour schemes (wireframes, such as complementary, split-complementary, etc.) is the most common basis for the underlying order.

2.9.8 The readability of text

The coloured items in an interface almost always include some text and it is therefore necessary to consider how the selection of colours for the text and its background affects readability[30]. There has been extensive research on the factors influencing readability, primarily motivated by both ergonomic and safety concerns. Readability is critical to the usability of an interface. It is therefore important to determine the dominant factors affecting readability of textual interface items, to enable the algorithmic assessment of the readability of text in an automatically derived colour scheme.

It seems reasonable to assume that if text is a very different colour from its background (i.e. there is a large colour space separation) then the text would be easily readable. However, this is not always the case, as can readily be demonstrated with saturated green text on a red background (fig. 2.28) which is difficult to read due to an unpleasant shimmering effect. The reason for this is detailed in section 2.2.6.

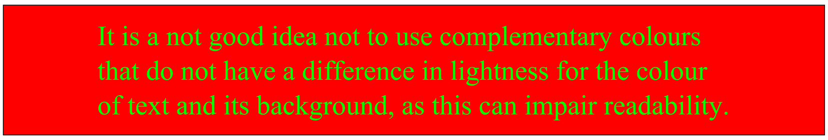


Figure 2.28: The use of complementary colours for text and its background can impair readability.

For an item to be visible against its background, there must be contrast, either chromatic contrast, lightness contrast, or both. However, for text (as is evident from fig. 2.28), chromatic contrast alone is not a reliable predictor of readability. As will be seen, lightness contrast gives more reliable indication.

Text contrast and its polarity

Light–dark contrast may be obtained in two ways: the text may be displayed as dark text on a light background (as on paper), or light text on a dark background, as was common in the early text-only visual display terminals that used green, orange or white text on a dark screen background. Dark text on a light background has become known as a *positive polarity text*, and light text on a dark background is known as *negative polarity text*. It is necessary to take care when comparing results, as the nomenclature is inconsistent. Dark text is sometimes referred to as positive text or text with a positive polarity (Fukuzumi and Hayashi, 1989; Buchner and Baumgartner, 2007; Fukuzumi et al., 1998) and negative polarity (Snyder et al., 1995; Scharff and Ahumada, 2005).

The contrast can be measured using either the *luminance modulation*, or the *luminance contrast ratio*:

$$L_m = \frac{L_l - L_d}{L_l + L_d} \tag{2.12}$$

$$L_c = \frac{L_l}{L_d} \tag{2.13}$$

[30] The text discussed in this section is assumed to be information-carrying, not decorative.

where:

L_l	–	lightness of the lighter item.
L_d	–	lightness of the darker item.
L_m	–	luminance modulation.
L_c	–	luminance contrast ratio.

A small factor corresponding to ambient flare[31] can be added to each of L_l and L_d to ensure that neither is zero.

The research on whether the polarity of text makes a difference is mixed. Several studies have found a significant advantage in using text with a positive polarity (dark text) and others have found little difference. Wang and Chen (2000) found that increasing contrast increases performance, but polarity was *not* significant. However, this trial did not test the readability of actual text. Instead a graphic was used: the Landolt-C visual acuity test symbol (like a circular small letter c). This was done to remove the actual text and its characteristics as a possible source of bias. If other factors (e.g. word shape recognition) affect readability and success is not solely related to perception and visual acuity, the lack of real text may explain the discrepancy in the results. For both polarities, increasing the luminance contrast correlates well with increasing readability (Lin, 2003, 2005; Fukuzumi et al., 1998).

The results of the recent study by Buchner and Baumgartner (2007) are unequivocal in showing that text with a positive polarity (dark text on a light background) is more readable under all conditions of ambient lighting, and the chromaticity contrast alone is not sufficient for comfortably readable text. This agrees with earlier studies by Snyder et al. (1995) and Scharff and Ahumada (2005). Gradisar et al. (2007), while noting interactions with chromatic contrast, also found that, overall, text with a positive polarity contrast is more readable.

The importance of luminance contrast is well summarised by Ojanpaa and Nasanen (2003): *"reading rates decreased with decreasing luminance contrast. Thus, moderate or even high colour contrast does not guarantee quick visual perception, if the luminance contrast between characters and background is small. . . . Therefore, in user interfaces, good visibility of alphanumeric information requires clear luminance (brightness) difference between foreground and background."*.

The contrast required for readability

A contrast ratio of 5.6 ($L_m = 0.7$) is regarded as acceptable by ISO 9241 for colour displays[32] according to Smith (1996), cited in Bangor (1998). The distinguishability of fine detail is primarily dependent on lightness difference, as noted by Klassen et al. (1998, p27): *"we know that chromatic differences are more important in large regions than in small features and high frequency regions, where lightness differences are more important"*. The W3 consortium, responsible for the World Wide Web standards,

[31] Flare – light reflected from the surface of the display
[32] For monochrome displays, a lower contrast ratio of 2.3 is acceptable ($L_m = 0.4$).

reflect this finding in the web accessibility guidelines where a contrast ratio of 5:1 is recommended for small text. This is relaxed to 3:1 for large (18pt or greater) text.

In experiments under controlled conditions (Zuffi et al., 2006) and later as a web-based experiment using a much larger sample size (Zuffi et al., 2007), it was found that a difference in CIE lightness difference of ΔL of 25–30[33] was sufficient for readability, irrespective of polarity. If $L_d = 10$ then $L_l = 40$ would correspond to a luminance modulation, $L_m = 0.6$ or a contrast ratio of $L_c = 4:1$, comparable to the W3 recommendations. These studies are particularly significant here as they are oriented towards the readability of text on web-pages and the experiments included text down to 8pt in size. In printed documents, 8pt is quite readable but not common, with 10–12pt being more usual. However, the limited screen real-estate available on web sites and the desire of web site developers to have as much information visible "above the fold" (without requiring users to scroll), has lead to different design criteria. On web sites, as confirmed by Zuffi et al., the use of 8pt text is quite common.

Readers with impaired vision can find the readability of smaller fonts to be especially difficult. Bangor (1998) recommends a contrast ratio of at least 18:1 for text smaller than 18pt in size, especially in the case of negative polarity text.

Informal experiments by the author, using medium saturation complementary hues at 30° hue increments in CIELAB space found an average ΔL of 48 was necessary for easy reading of 8pt type for both polarities. The test values were centred around $L = 50$ with the achromatic colourings requiring slightly greater difference than all rotations with complementary chromatic contrast. The measured values give contrast measure of $L_c = 2.85$ and $L_m = 0.48$, in moderate agreement with the W3 values. The values are somewhat larger than those of Zuffi et al. The tests were conducted on a 20 inch monitor at its native resolution of 1600×1200, resulting in slightly higher dots-per-inch (101) than the 96dpi of Zuffi et al.. The higher resolution results in slightly smaller characters and as noted by Bangor (1998), readability drops off very quickly with text below 18 minutes of arc. As the smallest text in the text was approximately 5 minutes of arc, it is not surprising that slightly enhanced contrast was found desirable.

It might be expected that the font used to render the text could affect readability. However, Garcia and Caldera (1996) report that the font used has less effect than text colour on the time taken to read text.

It is clear that with small text at moderate lightness contrast, positive polarity text will be more readable, but given sufficient lightness contrast, the readability of text of either polarity and any hue can be guaranteed. This is useful as it means that lightness difference alone can be used to ensure that the text in an automatically derived colour scheme will be readable.

[33] ΔL ranges from 0 – 100 (black – white).

2.10 Computer-based tools for colour scheme design

The previous sections have demonstrated that colour scheme design is important in user interface design and that there are methods that can be used to select colour schemes likely to have broad appeal. This section will survey the capabilities of existing software for selecting colours and creating colour schemes oriented towards design, with a particular emphasis on user interface colour schemes.

2.10.1 General purpose colour selectors

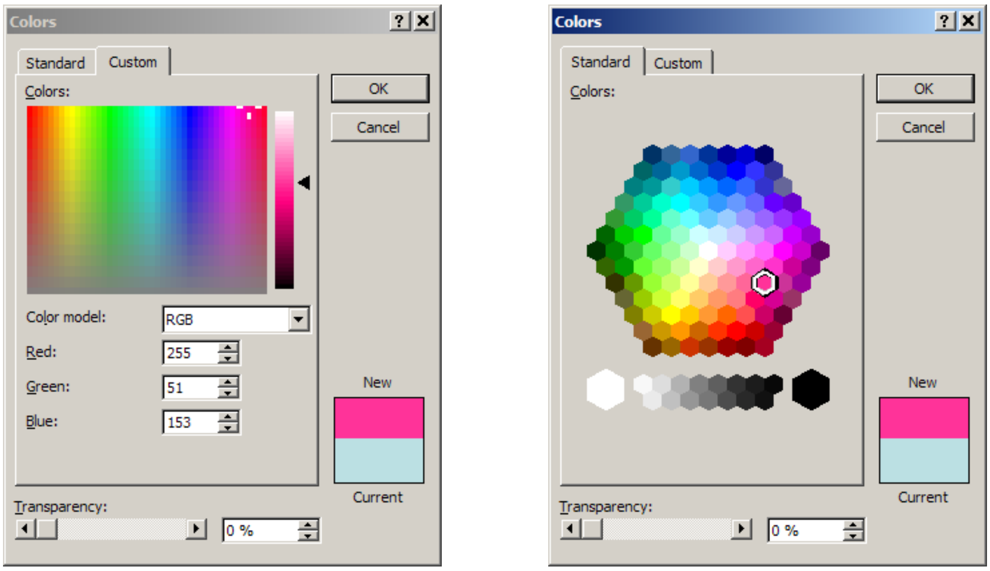


Figure 2.29: Two colour selectors typical of those in current use, both from Microsoft *PowerPoint 2003*™. On the left, a flattened HSL representation, and the right, a form resembling a colour wheel. Neither supports selecting sets of colours, and both are modal, forcing the choice of colours out of context. It is very common for repeated invocations of either dialog to be necessary, as the new colour (visible in the swatch at bottom right of the dialog) can appear different when viewed in context, the visual impression being affected by the area of the colour and by nearby coloured objects.

The use of colour in the desktop computing environment has changed significantly over the last two decades. Any modern computer is able to display millions of colours, and software for the creation of documents and graphics is routinely able to let the user select the colour of any document item. However, the facilities offered by most colour selectors are oriented towards single colour selection; as illustrated by the *PowerPoint*™ colour selectors shown in figure 2.29, there is little to aid the user in the selection of harmonious colours other than presenting differing views of the colour space. Despite research that indicates that immediate visual feedback improves user performance when choosing colours (Douglas and Kirkpatrick, 1996, 1999), such a facility is rarely available. Almost all colour selectors are modal, forcing the user to choose colours out of context, although this is starting to change, as illustrated by the non-modal colour selector from the

InkScape drawing tool[34] which is shown in figure 2.30.

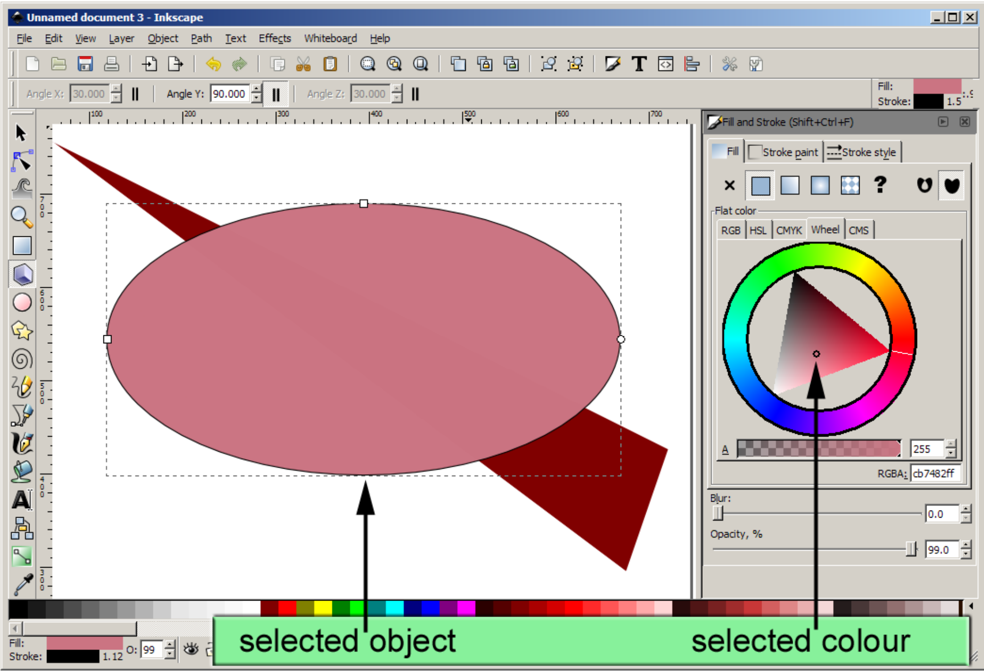


Figure 2.30: One of the non-modal colour selectors available in the InkScape drawing tool. The combined colour wheel and hue slice colour selector on the right enables all tints and shades of the selected hue to be seen simultaneously. Dragging the little circle (the selected colour) within the triangular area immediately updates the colour of the selected object. The colours are therefore seen in context, which helps to avoid colour selection errors.

One might expect that the colour selectors in programs oriented towards graphic design and the visual arts would be significantly more sophisticated than those general-purpose selectors surveyed above. This is indeed the case, but the additional facilities are generally more oriented toward professional designers who know how to create and use palettes appropriately, rather than features that would directly aid naïve users when colouring interfaces. For example, products from Adobe's *Creative Suite*™ range include two features particularly relevant to colour selection. The first is the ability to work in colour spaces suited to the destination domain, such as specific RGB colour spaces or CIELAB in Adobe *Photoshop*™, when manipulating images for on-screen use, or CMYK[35] for printed images. The spot colour selector is enhanced (fig. 2.31 left) and able to show views of colour space in ways that make it easier to find ordered sets of colours. Professional designers will often use multiple palettes within a single document, and so sophisticated palette management is supported (fig. 2.31 rightmost image). New palettes can be created by specifying a key colour, with the remaining colours calculated using pre-specified relationships with the key colour. The traditional colour relationships, such as complementary and split-complementary are included,

[34] `http://www.inkscape.org`, accessed May 21, 2010.

[35] CMYK – Cyan, Magenta, Yellow and blacK, the subtractive primaries used in printers.

along with many others oriented to specific colour associations, such as "ice-cream" and "corporate". The exchange of palettes using either palette files or Internet-based resources is also supported. For example, Adobe *Illustrator*™ supports the browsing and the importing of user-contributed palettes from Adobe's *Kuler*™ website[36] (Adobe Systems, 2009). Once a palette has been imported, the new colours can be easily applied to an open document.

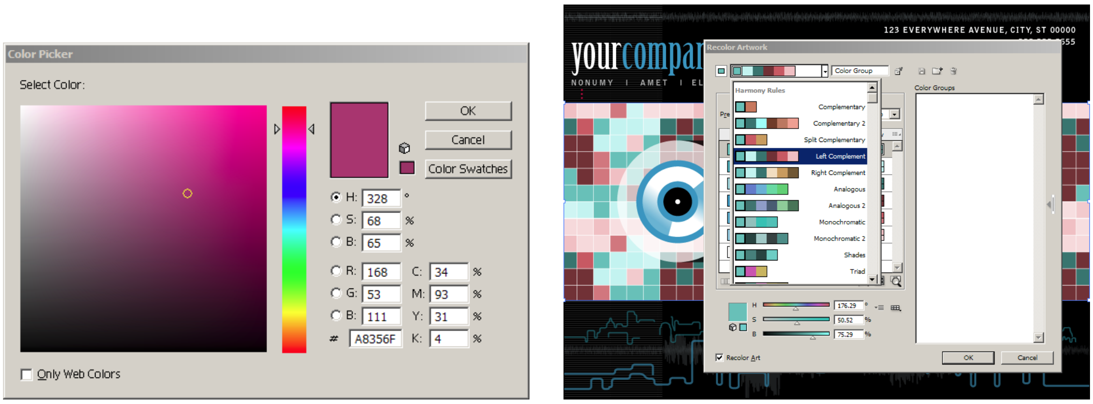


Figure 2.31: Two colour selectors from the Adobe *Illustrator CS4*™ package illustrating the features offered by professional graphical design tools. On the left, the contents of the thin vertical rainbow-coloured area can be set to show all variations of a particular colour property. The selector on the left displays all the hues in the vertical strip, with the larger area showing the other two colour space dimensions (for the example shown, lightness vertically and saturation horizontally). On the right, an illustration of the palette management facilities. The multi-coloured grid on the web page can be instantly recoloured by clicking on any of the palette sets.

These palettes are excellent starting points for artistically-oriented users, but do not directly address the needs of unskilled users. Programs like *Illustrator*™ and *Photoshop*™ are intended for use by professional designers who will use the palettes as starting points and modify the colours to incorporate any aesthetic or domain-oriented requirements that they deem appropriate.

2.10.2 Template-based colour scheme design

Users whose specialty is not graphic design would probably prefer a package intended for a non-specialists, such as Microsoft *PowerPoint*™. The colouring facilities offered by this program, probably the most widely-used presentation design tool, are based on templates, offering the ability to vary the presentation colours as a group, or altering the colour of individual elements (for details see Moretti and Lyons (2002)). The Microsoft *Frontpage*™ website design tool has colour selection features that are based on templates in much the same manner as those of PowerPoint, but augmented to cater for web-specific facilities such as the different colour of visited links or graphical elements used as hyperlinks (fig. 2.32). Template-based systems can work well, as

[36] `http://kuler.adobe.com/` - accessed May 21, 2010.

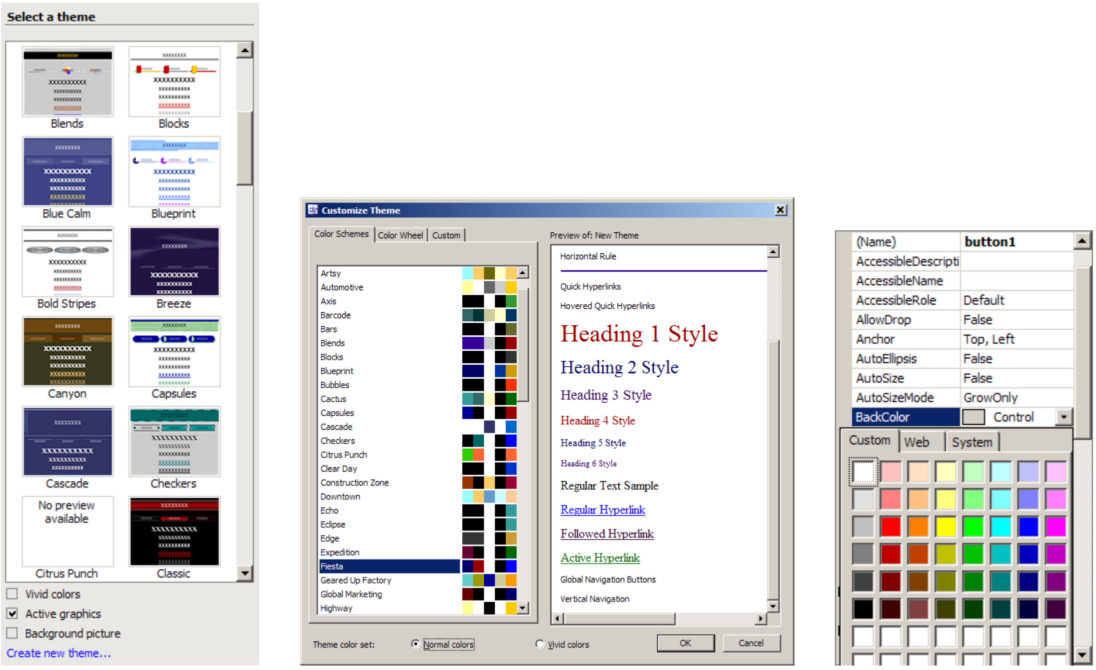


Figure 2.32: Left and centre: the template-oriented approach to colour scheme management from Microsoft's web design tool *Frontpage 2003*™. The facilities in *PowerPoint 2003*™ are similar. Complete schemes can be changed or recoloured in toto, but only with the indicated coloured elements. Right: the facilities offered to GUI application developers in Microsoft *Visual Studio Express 2008*™. The options, unless altered programmatically, are limited to choosing from the limited palette shown. Microsoft products are illustrated as they are very widely used, and their colour selection facilities are typical of those found in products from other manufacturers.

long as the design is limited to colouring the elements within the predefined templates. However, such a constrained approach does not allow the flexibility to colour user interfaces or websites with an indeterminate number of colourable elements.

Software design environments intended to create runnable desktop applications have strong support for programming features (popup help, program breakpoints and debugging support, refactoring etc.), but support for the creation of user interface colour schemes is fairly rudimentary. This is illustrated by the limited colour selection features (fig. 2.32) in the otherwise sophisticated *Visual Studio Express 2008*™[37] IDE for C#. This is not surprising, as the typical method of including user-configurable colour scheme – templates – does not generalise well. The range of possible interfaces is so great that it is impossible to create a generally applicable approach based on fixed templates.

[37] The Visual Studio suite (there are different versions) is Microsoft's multi-language software development environment.

2.11 Automating colour scheme design

The design process is clearly outlined by Eckert et al. (1999), who note the iterative nature of creative design and the advantages of generative systems (those that create designs). They classify such systems into "standalone" generative systems, those that are strongly biased towards a particular type of design, and more open-ended generative systems that allow the incorporation of user constraints and iterate through the stages of: *"problem specification, design generation and design evaluation"*. The open-ended option allows for the development from partial specifications and more closely mirrors the design iteration process used by human designers.

Relating this to colour scheme design: the template-based approach corresponds to the standalone categorisation – a predefined method achieves a well-defined result, but with little flexibility. The open-ended method can work with fewer initial constraints, but based on whatever constraints did exist, it would create a set of conforming schemes. The constraints could then be modified or redefined and a new set of potential solutions generated. The open-ended method is significantly more flexible. It requires fewer initial constraints, and is therefore capable of a more extensive search of the solution space, without precluding the inclusion of tightly defined constraints if desired.

Eckert et al. state *"for most interesting classes of artifacts, the space of possible designs is immense. At any stage in the construction of a design, the vast majority of possible changes are either nonsensical or foolish. So to create a design that meets its designer's objectives, the generation process must be strongly directed"*, but then note that *too strongly* biasing the design process is not necessarily desirable either, as it can preclude the discovery of designs that do not conform to usually-used solutions (i.e. those in less frequently used or unexplored parts of the solution space). The suggested solution to this dichotomy is to allow the users to provide such biases interactively, rather than preprogramming such constraints into the system. Without a database of colour preferences (e.g. "young girls like pink and purple"), the user may choose to use such schemes if they are thought to be appropriate, but is also free to explore other less obvious combinations. This level of interaction is more suitable for experienced designers who are familiar with the appropriate schemes for differing domains. If such constraints or biases were optional, designers with less artistic ability or less time could use the stereotypical schemes and designers with more artistic ability or time could develop more subtle designs.

Eckert et al. notes that generative systems could be used to create designs that are, because of the computational complexity, impractical for human designers. For colour scheme design, this might be thought not to apply, as artistic human designers do not use complex computational methods for colour scheme design, but derive a scheme using a combination of intuition and experimentation. However, computational methods of colour scheme design are an attempt to generate results equivalent to those of human designers, not necessarily to mimic the design process itself. Therefore, using such a model, a generative system could create colour schemes at speeds that would be impossible for human designers, and that (depending on the completeness of the model) may include schemes of comparable quality.

Computer-aided design (CAD) tools, those that encapsulate domain knowledge and simplify the creation of complex artifacts, are used in virtually all fields. There are however, very few systems that would classify as generalised "CAD for colour". The generalised palette handling tools mentioned earlier are not domain-specific, and therefore their results must be used appropriately – design skill is needed in their application. However, if knowledge of the final intended domain is included in the design system, much more sophistication is possible. This can be illustrated by the ColouriZe system (Bailey et al., 2001, 2003) that aids in the creation of architectural colour schemes. The framework uses colour-related categorical specifications including: the use of named colours; the type of building; the intended use of the facility; whether the colour scheme is for interior or exterior use; and the style of the building by genre or period. This system can also include a psychological profile of the client: *"psychological profiling can be achieved by posing simple questions to obtain an outline of a person"* – Bailey et al. (2001). The intention is to alter the perceived dominance of the architect in the design process, and to alter the balance in favour of the client. With the inclusion of more domain knowledge into the design process, the more algorithmic (and often tedious) parts of the design process can be automated, and human input reserved for the more creative or subjective parts of the process.

There can be varying levels of assistance with colour scheme design in design applications. The aforementioned predesigned (template) schemes have limited versatility, and systems that aid the user by maintaining palettes of colour swatches, while more versatile, are better suited to experienced designers. The following sections will explore three more sophisticated techniques: automatic colour palette design, interactive design using colours in context, and automatic colour scheme design.

2.11.1 Automating colour palette design

The palette creation facilities inside Adobe *Illustrator*™ mentioned earlier illustrate the use of fixed colour relationships and a key colour to automate the creation of a palette. Adobe *Kuler*™ uses another method – deriving palettes from the colours within pictures or photographs. These are palettes either intended to reflect known and pleasing colour relationships, or to extract colours from appealing pictures.

If colour schemes are to be designed algorithmically, the use of perceptually uniform spaces will allow the relationships between colours to be largely independent of hue. An example of such a system is the "Functional Colour Selection System" of Beretta (1994), who notes the difficulty caused by the uncoordinated selection of individual colours when a coordinated group is required, and describes a method using wireframes and fixed relationships in CIELAB space to create palettes based on a specified key colour and lightness or chroma variations. The intended users are those wanting to choose harmonious colour for use in presentations and graphic design. After removing any unwanted colours, the palettes may be saved to a database. The publication is in the form of a patent so, while reference is made to an implementation, it is not described. This work follows the earlier "Reference Colour Selection System" (Beretta, 1993) that allows the creation of palettes, and includes the ability to add colours by taking pairs of colours from the existing palette and producing interpolations between them. The system of Luke and Luke (2000) also uses interpolation in perceptually uniform space to derive palettes of harmonious colours. Luke and Luke state: *"when the positions of two points within a uniform three dimensional space are known, this invention finds all the planes in the space that pass through an axis determined by those two points."*. This too is a patent specification, so while the method is outlined, no implementation is described. The palettes created by Beretta and Luke and Luke are not constrained by any application-specific requirements. However, if the application area is known, the creation process can select the colours to suit the intended application area – for example, in data visualisation.

Palettes for data visualisation

The "design of colour schemes" is often interpreted to mean the design of palettes, and there are a wide variety of tools available. As previously mentioned, the design of palettes differs from the design of colour schemes, but there are commonalities as can be seen by comparing the design of palettes for use in data visualisation with the design of colour schemes for a user interface. For both, discriminability can be important, and a pleasing appearance is desirable. The need for such tools is illustrated clearly by Tufte (1997) in his scathing criticism of the naïve use of colours in visualisation schemes. Commenting on an oceanographic map with depth encoded using ten high-saturation colours spanning the visible hues, he states:

> *"in ghastly contrast . . . a rainbow encodes depth. Although often found in scientific publications, such a visually naïve color-scale would be laughed right out of the field (or ocean) of cartography. These aggressive colours, so unnatural and unquantitative, render the map incoherent" – Tufte (1997).*

He is promoting the *smallest effective difference* design principle: *"make all visual distinctions as subtle as possible, but still clear and effective"*, in order to make the best use of the range of available colours and minimise visual clutter.

Data visualisation systems are primarily presentation frameworks, and while interaction may be possible, it is not the primary purpose of the system. The visualisation within a data visualisation system has similarities to web/application user interfaces; both have foreground and background elements, and text, if present, must be readable. While there may be discretion in the colouring, the form of the display is not under the developer's control, as it related to some aspect of the data being visualised. Therefore, the designer has less control in achieving a pleasing appearance than when colouring an interface. This is especially so if the visualisation is dynamic, with newly arriving data altering the display and the balance between the coloured regions. Therefore, data visualisation systems do not usually have a predictable balance or partitioning of the coloured regions.

The regions that require colouring in a data visualisation system are task dependent. Commonly, the regions are either predefined or determined from the data – the designer does not have the flexibility to group regions – and distinguishability between regions is usually necessary. While the smallest effective difference principle could be used, it may not be appropriate. The placement of colours is not necessarily well-ordered, so larger differences may be desirable. This would allow a clear correspondence between the colours in the legend and colours in the visualisation, or make it obvious if widely separated regions are (or are not) the same colour.

In data visualisations, the areas of colour and their placement are determined by the data, and are therefore apt to change as the user alters the visualisation parameters or as new data arrives. This limits the ability of an automatic colouring system to alter the visualisation colours, as doing so may well result in the disturbing effect of the colour scheme changing with the data.

The following sampling of the tools for creating visualisation palettes illustrates some of the approaches used:

- Healey (1996) describes a system in which particular attention is paid to ensure visually distinctive categories, so that no anomalies are inadvertently missed when searching medical images. The approach combines perceptually-uniform colour space distance, and colour categories derived from the Munsell colour names to create visualisation palettes. The CIELUV space is used and *"we found that an isoluminant slice through a monitor's gamut yielded up to seven different colours, any one of which could be rapidly and accurately detected, even in the presence of all the others"*.

- The system of Cooper and Kamei 1999; 2001; 2002 aids novice designers when colouring visualisations, by deriving palettes with balanced conspicuity (conspicuousness). Conspicuity is shown to have a significant relationship with ease-of-use. The system uses experimental data to train a neural network that produces a suggested palette of colours that the designer is free to use in the visualisation or not.

- Ihaka (2002) uses ideas based on Munsell's balance to create palettes for use in visualisations. The system creates palettes of colours in the CIELUV space balancing the colours either around mid-grey or offset balance points to yield palettes from either the warm or cool sides of the colour wheel. The palettes are pleasant, and the colouring unambiguous. The number of colours shown is small (four and six colours), and while the equi-spacing of colours could be used to create larger palettes, the upper limit of easily distinguishable colours is less clear. The author comments on the desirability, in some contexts, of manually reordering the colours, so that common associations are preserved.

- Campadelli et al. (2001) detail a method of allowing the user to choose sets of high contrast colours from the Munsell colour space aided by an algorithm named the "Minimum of the Sum of the Inverses of the Dissimilarity" with colour difference measured after converting the Munsell colours into the CIELAB colour space. The possibility of forcing certain colours to be present in the resultant set is included, along with the creation of colour sets based on a triangle or square in Munsell colour space, in the same manner as a wireframe on a colour wheel. In experimental trials, manually created schemes had poorer distinguishability – the colours chosen had less than half the CIELAB colour space separation of the schemes produced by the automated system.

Visualisation systems have a number of regions to be coloured, a requirement for these regions to be visually distinct (possibly with a user-imposed ordering), and often the perceptual distance between the colours is related to some measure or difference in the input data. The use of colour wheels and perceptually uniform spaces for creating both visualisation palettes and colour schemes for various purposes could give the impression that there is little difference between visualisation palettes and harmonious colour schemes, but this is not so. Of the following three points identified as useful in deriving harmonious colour schemes, only the first also applies to visualisation palettes:

perceptible order in the colouring: ideally the viewer will be focussing on the information from the visualisation itself, not the colours used, but if the colours have been chosen from a path in the colour space, this can provide a pleasing sense of order. Rogowitz et al. (1999) explores in detail which colour space trajectories are most suited to visualisation palettes. Perceptible ordered colours are therefore both possible and, for aesthetic purposes, desirable.

the use of complementary colours or related colours: this heuristic is intended to remove ambiguity in a harmonious colour scheme, to avoid the need to consider the question "are these colours/regions related?". By contrast, the intention with data visualisation is to pick clearly distinguishable sets of colours, and therefore fairly large separations in colour space would be the norm. While the use of complementary colours may coincidentally arise as a result of maximising colour distance, this is not related to any desire for colour harmony. Therefore, neither related colours (which are not necessarily distinguishable) nor the use of complementary colours, are particularly applicable to data visualisation colour schemes.

the balance of colour strength: this is only possible when the data, and therefore the areas of the regions to be coloured, are known. For dynamic visualisations, the colours could change in lightness and saturation, and possibly hue, as the input data varies, an effect that would be visually disturbing. Balancing colour strength is possible only with a static data set, and therefore cannot be generally required.

It is therefore clear that although the selection of colour schemes for data visualisation and the design of harmonious colour schemes have factors in common, they are distinct problems. For more detail on the use of colour in visualisation systems, see Rogowitz and Treinish (1998) and Silva et al. (2007). The wide diversity of design possible in visualisations, graphics and user interfaces is clearly illustrated in Woolman (2002), and the perceptual considerations are considered in detail in Tufte (1990) and Tufte (2001).

2.11.2 Interactive design using colours in context

Between the design of palettes without context and those that automatically create user interface colour schemes, there are systems that both allow the creation of palettes and provide a way for the user to view these palettes in context, i.e. as colour schemes. Bauersfeld and Slater (1991) describe work allowing the visualisation of 3D perceptual colour space and its use for colour selection, and the use of resizable draggable swatches to visualise the impact of coloured regions in context. A more recent and more extensive exploratory framework is that of Meier et al. (2004), which is designed to allow the exploration of answers to questions such as[38]:

- *"what goes with this color?"*
- *"what is a good background or text color?"*
- *"what are two (or several) colors that look well together?"*
- *"how can I get a color that is a blend of this green and blue?"*
- *"how would my design look if I added some purple to it?"*
- *"how would my composition look if all the colors were more subdued or lighter?"*
- *"these colors are close to what I want; how can I get some palettes similar to this one?"*
- *"can I find a color like brown by searching for it by name?"*
- *"how can I arrange my swatches so that all the reds are near each other or all the dark colors are together?"*

Support is provided for: the creation and management of sets of palette swatches; the creation of gradients (interpolated colours); palettes created using a key colour and a wireframe (called "colour harmony rules"); grouping related palettes; and a "palette breeder" that combines two "parent" palettes to produce derived palettes, combining features from the parents and incorporating random changes. The palette breeder uses ideas from genetic algorithms (GAs, which will be introduced in sec. 2.11.3). Meier et al. (2004) state: *"the breeder, although inspired by genetic algorithms, isn't a genetic algorithm because it has no formal evaluation function for judging the palettes generated"* and of the schemes produced by the breeder: *"the variations aren't ones that typical users would have created on their own, but they offer fresh color ideas based on existing palettes. This is more useful than completely random palettes, which usually aren't aesthetically compelling"*.

This system is positioned between those that create palettes without considering the final composition, and those that use information from the composition during the design of the scheme. Effectively, it is an exploratory colour workbench. A basic drawing tool is included to allow the creation of simple "thumbnail" compositions that include text and shapes. It appears that composition elements can be linked to palette items: the mechanism for this is not detailed, but the association is one-way – colours in the composition are affected by colours in the palette – which allows the composition to be recoloured immediately when the palette changes, but the properties of composition elements are not used in deriving palette colours. This tool would be an

[38] Quoted from Meier et al. (2004)

ideal exploratory framework for graphics design professionals, but like the swatch-based tools incorporated in high-end graphics tools, the user needs to understand principles of graphic design to use the tool effectively.

2.11.3 Automated generation of interface colour schemes

The derivation of harmonious untargetted palettes[39] and palettes for use within data visualisation can be automated. The design of "colour schemes" is more complicated as the colour choice is influenced by the object being coloured, and the design of user interface colour schemes is even more complicated, as usability constraints must be considered. There are guidelines on colouring user interfaces (Murch, 1984; Travis, 1991; Wright et al., 2000; Shneiderman et al., 2009), but most are lists of points to consider, not methods of choosing colours or schemes, and there are far fewer implemented systems specifically oriented towards the creation of GUI colour schemes.

An expert system approach to colour scheme generation

The "ACE colour expert system" (Meier, 1988) is an early example of a method of automating the design of user interface colour schemes. The interface being coloured was that used on the XEROX Star and early models of the Apple *Macintosh*™. This system clearly demonstrates how the underlying model can greatly affect the complexity of the system, the quality of the resultant schemes, and the generalisability of the approach.

The interface style at that time (1988) consisted of the desktop with its icons and the top level menu, and an application-specific toolbar and main window containing text and/or graphics. The partitioning of windows into multiple sub-panels, each containing more controls, was less common then than it is now. Consequently, the interface style was simpler, consisting of an application window, the top-level menu-bar and icons on the desktop. Nevertheless, the inclusion of both the menu and desktop into the controllable items provides significant scope for colour scheme design with the following items being controllable: window background, window text; grids; graphical shapes; scroll-bars; menu items (selected and not); desktop icons (foreground and background; selected and not); and the application toolbar (called the "permanent menu").

ACE was implemented as a forward-chaining production-rule system using the OPS5 expert-system framework. The knowledge base contained:

- *rules about single item colouring and relationships* e.g. to ensure distinguishability of layered interface elements;
- *rules about harmonious colour relationships*;
- *control knowledge*, i.e. rules that encode methods for proposing constraints and colours.

[39] those created using colour harmony rules without considering the final use.

The execution of these rules (170 in the prototype system) results in a suggested set of colours.

Initially, the user describes the properties of new objects and their relationships to existing objects using the predicates in-front-of, behind, beside, and on-screen at same time (or not). From this information the ACE system creates:

functional constraints: these relate to single objects and constrain the possible range of colours for that object, for example: *"the desktop is static, in the background, and large in size ... therefore, according to our color rules, it should be a dark, desaturated colour"*. These constraints, once imposed, cannot be retracted and act to reduce the range of possible colours.

relational constraints: these apply between pairs of objects. e.g. a constraint may be asserted that requires a specified contrast between pairs of adjacent items. The constraints can act either to force the colours apart, or to be identical. Unlike the functional constraints above, relational constraints are not permanent (i.e. they can be retracted).

global constraints: constraints that would affect the overall colouring are posited but not implemented. As an illustration, the addition of a rule to promote the use of desaturated colour schemes for interface schemes likely to be used for extended periods is given.

colour assignment: items are assigned to one of five categories, but two are sufficient here. In outline, large and background areas are placed in one category, and small foreground items in another. Colouring proceeds by assigning colours to the large item category, selecting items in random order, until all within the current category have been assigned. Then items in the other category are coloured. This order of colour assignment therefore allocates the larger, more visually imposing objects first, as these are the objects that act as backgrounds to, and therefore interact with, a larger number of foreground objects.

There are two phases to the colour assignment process:

collecting proposals: proposed colours or constraints on the brightness, saturation and hue: *"ACE tries to find a hue relationship pair in the colour pool that matches the proposing object's hue, the adjacent and screen contrast values, the relative brightness direction, and that also has the highest attractiveness value"*. If a hue cannot be found, constraints are relaxed, until, eventually, grey is used.

grading: once all the proposals are in, a grading process begins. This involves each proposer assessing all the other proposals, to determine how far the other proposed colours are from the proposer's ideal colour. This difference is used to alter the grade of the other proposals: *"an object with a large weight has the effect of hurting every other candidate's grade while an object with a small weight does not"*.

> All the proposed object colours are ranked by grade, with the winning colour being assigned to the object under consideration. This process continues until colours have been assigned to all objects.

At the end of this process, the result is a single colour scheme, containing a colour for each interface element. The scheme is displayed in numeric form and could be taken and then used as the basis for further development, but such exploration was not part of the ACE system.

The system was an innovative attempt to encode the "soft knowledge" of colour scheme design into production rules. In the absence of a general colour harmony assessment function, this assessment is done at a very detailed level, encoding relationships between specific pairs of colours:

> *"we tried to discover a general relation between any two colors in a three-dimensional color space that would show whether the two colors harmonized or contrasted and how attractive they appeared together. This involved extending the one dimensional (hue only) harmony/contrast relation of the color wheel to three dimensions and adding attractiveness relations as well. Unfortunately, we were unable to find any general relations, so we selected a discrete set of three-dimensional colors and explicitly tabulated the relations between them based on aesthetic judgment"– Meier (1988).*

The ability to add new rules to enable the creation of new styles of colour scheme is appealing. However, in an OPS5 system, the firing of rules is based on the currently applicable situation, not any fixed order of execution. The rule-based structure fragments the approach to colour scheme creation into multiple rules, each applicable in particular circumstances. This has been identified by the authors as a shortcoming: *"another drawback to the way that ACE selects colors is that both selections and evaluations are performed pair-wise, but the effect of a group of colors is as a whole. We might be able to devise color relations between three, and maybe even four or five colors, but more than this would be extremely difficult."*

The schemes produced are stated to be better than random schemes, and those created by naïve programmers, but no results from user trials are reported.

It is clearly stated that the created schemes benefit from further refinement. How this could be accomplished is less clear:

> *"in general, the visual effect of color is hard to encode. Color relationships are very subtle and a small change can make a big difference. A designer might spend several hours tweaking the colors of a user interface. . . . There seems to be a limitation on the aesthetic quality of ACE's output; ACE can select reasonably attractive colors, but it probably will not be able to perform the same fine adjustments that a human expert can. We may be able to incorporate some of these fine adjustments, but first we need to examine the ways that different experts solve the problem"[40] – Meier (1988).*

[40] Meier's later work, as discussed in section 2.11.2, moved away from a fully automated approach towards one that includes more user interaction.

The rule-based approach, in performing an ordered selection of colors for individual items, suffers from a lack of overall perspective. The addition of more rules, in an attempt to encode expert knowledge, cannot be guaranteed to help in gaining such a perspective. The development of expert system rule sets is very time-consuming and, by its nature, difficult. The knowledge elicitation phase – getting experts to explain the steps they use to arrive at a solution – is prone to error and inconsistencies, and the rules are frequently incomplete. Checking the correctness of the rule base becomes more difficult as the number of rules increases, and the presence of contradictory rules, or rules that will fire in inappropriate circumstances, is less evident. This is especially so in the aesthetic or artistic fields, as there is not one "right" way, but many possible arrangements, each of which may be internally coherent, but not amenable to decomposition into rules that may be applied in a different context. The brittleness of the expert system approach is well known (Giarrantano and Riley, 1998; Fogel, 2005) and given the difficulty of elucidating artistic design principles, the reliable generation of harmonious colour schemes using an expert system appears problematic.

The dynamic recolouring of desktop windows

MacIntyre (1991) describes a system designed to colour multiple related windows in an aesthetically pleasing way, providing usable defaults, allowing the user to specify colouring constraints and gradually customise the results, with the author noting that *"typical users make far better critics than designers"*.

The system is modelled on a dynamic window positioning system that reorganises windows while the system is being used (Schlueter, 1990). Since colours can change while the system is in use, to avoid disorienting the user, the authors stress the need to have the colours change gradually and be obviously related to actions by the user.

To colour a scheme, the user selects two colours and the prototypical colour scheme (wireframe), and the system chooses the remaining colours: *"each colour can be generated to fall randomly within the colour scheme, subject to whatever other constraints act on it"*. The author notes that random colour assignment is only viable when the user can easily discard one set of generated colours and create another. It is also possible to modify individual colours, or to lock the colour of one or more items and have the system create a new scheme incorporating these fixed colours. Colours can be specified by name (e.g. red), which gives the constraint solver some flexibility in choice of the actual colour used as there are many reds, while still incorporating the user's preferences into newly generated colour schemes.

The system implements dynamic window colouring, and in order to create reasonable schemes prior to any user specifications, the following design criteria are used[41]:

- *"border colour. Any hue, a saturation of 25%, a lightness of 50%."*
- *"border text. Absolute colour value of black."*
- *"window interior. Same hue and saturation as the border, a lightness of 90%".*

[41] Quoted from MacIntyre (1991)

- *"window text. Absolute colour value of black."*

Once a hue is specified, these constraints will produce pleasant, but bland, schemes that are suitable as default colour schemes, and can (with the specification of additional constraints) be improved. The colour scheme models are monochromatic, complementary (with the fixed relationships between window border and background colours noted above), and analogous.

The system colours multiple windows to indicate whether or not they belong to the same application or not, by using the same background for related windows. The implemented system uses a *distributed jostling model* (Schlueter, 1990) in which the constraint solver is modelled as a dynamical system with attraction and repulsion between the coloured objects, using colour distances measured in the perceptually-uniform CIELUV space. The area of objects is used during the optimisation to indicate the "mass" of the objects, with those of larger mass being less easily jostled and therefore slower to change colour. Therefore, while area is incorporated into the model, it is not (intentionally) related to any aspect of a colour harmony model.

The system is oriented towards colouring multiple windows to indicate the relationships between the currently visible windows in an aesthetically pleasing way. It does not attempt to colour the contents of the windows themselves, with the author stating: *"one class of windows that have seemingly been ignored throughout this thesis are those which contain realistic and/or complicated interiors"*. This is exactly the problem that the current thesis is attempting to address.

Automatic colouring of multimedia presentations

The system described in outline by Nack et al. (2003), and in more detail by Manniesing (2003), is the colour design module forming part of the *Cuypers* project. The aim of the project is to automate the creation of hypermedia presentations from multimedia data for use in a museum of fine art.

In the Cuypers framework, the structure, style and marked-up content are managed separately. Manniesing details the design and implementation of a module to colour a presentation: the spatial layout and temporal structure have already been defined. Given a presentation, a set of constraints and desirable goals, the colour design module will automatically colour the presentation.

The system differentiates between foreground and background areas and can alter the priority given to each. The colouring process depends on whether the user indicates that precedence should be given to the form of the interface, to its functional aspects, or a blend of the two. Differing colouring strategies are used in each case, depending on: *"whether the emphasis of the colour design is oriented towards the functionality of the information elements (foreground and structure) or rather on a look and feel, emotionally oriented presentation (emphasis on background and contrast)"* (Manniesing, 2003). The colouring method is based on rules, using details of the intended user group, the domain and the layout. It derives a list of the areas on the interface and the sub-elements contained within each area..

A hierarchical decomposition of the presentation is used to derive a list of items to be coloured, and using a list of colours appropriate to the domain (of six or fewer colours, to limit the visual complexity), the colour module attempts to find appropriate colours using predefined colour harmony schemata (wireframes). These are tried in order of ascending complexity: achromatic first, then monochromatic, complementary and finally split-complementary. The scheme used is determined automatically and depends on the number of colours required and the currently applicable constraints.

The HSL colour space is used as the basis for the colour harmony calculations. Legibility constraints are evaluated using multiple criteria to ensure a minimum standard is reached, and, if this is not possible, then black is used for the foreground text element and the system proceeds to colour the background. Requirements relating to state can be incorporated into the colouring. For example, hyperlinks can be either visited or unvisited. A suggested method of making such state-specific information visible without significantly affecting the colour scheme is to halve the saturation of visited links. An interesting aesthetic optimisation is the avoidance of pure black and white, by adding a small amount of colour to each. This gives a faint colour tint to the whites, and a hint of colour to the blacks, reducing the harsh contrast, and introduces an element of subtlety to the colouring.

This system is intended to create multimedia presentation interfaces automatically. The aim is to allow the content of multimedia databases (e.g. images of paintings, and commentary) to be displayed using a colour scheme that is appropriate and ensures readable text. This is not a tool intended for general user interface design. The existence of marked-up content is assumed, including lists of appropriate colours to be included in the presentation (domain colours), and the result is the generation of a single colour scheme. The colouring rules do not appear to take area into consideration when deciding on colours, with the result being that the schemes, while harmonious in hue, can be rather strongly coloured.

The underlying Cuypers system is impressive, as is the colouring module, but the need to produce a single colour scheme without human intervention, which is necessary for its intended use, means there can be occasions when it does not completely succeed. Given the complexity and subtlety of colour scheme design, this is not surprising. The Cuypers framework is intended to automatically produce acceptable and usable colour schemes for transient use, not ones that may be in front of a user for long periods of time, so an occasional non-optimal scheme is acceptable.

A different way of using rules is illustrated by the CRAFT system, which functions as an interactive design assistant (Rogowitz and Rabenhorst, 1993). It uses sets of design rules to constrain the colours available for colouring a window and its text. The colours available to a user for later selection are constrained by their earlier choices, excluding colours that would result in inharmonious or unreadable schemes. As the authors state: *"the user controls the selection of colors and fonts, but the system guides these choices through a system of perceptual rules. This approach can be described as a direct manipulation methodology supplemented with rules, or as an interactive expert system"*. Including the user in this manner avoids the difficulty of attempting to choose one particular set of colours. With the constrained selection method, the colours in the final scheme are those selected by the user. Whether the inability to access excluded parts of the colour space is a cause of frustration is not mentioned. To clarify what colours are available, multiple views of the 3D colour space can be shown simultaneously, allowing the usable and excluded colour ranges to be clearly seen. The colour space used was not perceptually uniform, although this was a planned enhancement. With the addition of additional rules, the creation of schemes with particular characteristics would be possible, such as those intended for users with colour vision deficiencies, or to satisfy particular usability, readability, or standards requirements. This system differs in approach from the ACE system in that, while it is rule-based, it does not attempt to decide a final scheme for the user; rather, it narrows the possible colours, but always leaves the user in control.

An alternative method of acting as a design assistant is to critique the designer's choices without constraining them. This is the approach used in the "eMMaC" system of Nakakoji et al. (1995), a tool designed to aid novice multimedia authors in the creation of presentations. The authors comment that the difficulty is not usually a lack of features in the multimedia preparation software. With novice authors, like non-colour-literate developers, it is more usually a lack of any clear goal. They note that much modern multimedia design software is so complex that, even if a particular function is known to exist, it may be extremely difficult to find. The system assists in three ways, providing information on colour combinations, colour balance, and colour associations.

The presentation used as the basis for the discussion had very few elements, and it was therefore sufficient for the colour critic to analyse the presentation using a colour histogram to find the most commonly used colours. The user is alerted when there is a significant mismatch between their colour use and that given by Munsell's law of colour harmony. Suggested recolourings of the less dominant elements are presented, along with messages relating to colour usage and associations. It is possible to specify the intended result using categories of theme colour and the thematic association of the presentation: *"color temperature (e.g., warm, cool), atmosphere (e.g., happy, sad, cheerful), types of titles (e.g. instruction, entertainment), target age (e.g. children, adults, elderly), and geographic and racial culture (e.g. Japan, Greece, USA)"* (Nakakoji et al., 1995). The user can make queries directly, e.g. to find appropriate colours for use in a given context.

Three factors are noted as causing particular difficulty for inexperienced designers[42]: *"users do not know about the existence of information; users do not know how to access information; and, users do not know when to use information"*. These are very similar to the factors that cause difficulties for non-specialists when creating colour schemes: non-specialists are frequently not aware of colour harmony heuristics; they don't know how to find *applicable* colour design information; and from the plethora of information they do find, it is unclear what is appropriate in a given context. It is therefore essential that any colour design environment intended to be used by novices remove superfluous detail (including choice), either by constraining choices so that major design errors

[42] quoted from Nakakoji et al. (1995) – additional explanatory text has been omitted.

cannot be made, or if errors are allowed, by drawing attention to them and suggesting a remedy. The latter option has the beneficial side-effect of educating the user.

Ou et al. (2008) present an early version of an exploratory tool that uses empirically-derived colour harmony rules (sec. 2.9.7). The work follows the partitioning of colour space into psychologically-meaningful dimensions, like that of Kobayashi's "Colour Image Scale" (Kobayashi, 1981, 1987). The scale used by Ou has dimensions warm–cool, heavy–light, and active–passive, and enables the use of linguistically appropriate categories in colour scheme design:

> *"This system allows the user to filter out an original, vast colour palette into a smaller amount of colours by specifying semantic feelings (e.g. warm/cool), and to create harmonious colour schemes by selecting colours recommended by the harmony model derived in this study. The software also allows the user to test colour schemes by presenting simulated coloured images of the design, together with quantitative evaluation (provided by the software) of selected colour schemes in terms of semantic associations and harmony." – Ou et al. (2008).*

The system uses a set of colour harmony rules derived from experiment that are significantly more restrictive than the colour harmony models discussed in sections 2.9.6 and 2.9.7. The more limited paths for choosing harmonious colours sets offer fewer possibilities for satisfying the competing constraints in a user interface colour scheme, although the resulting colour schemes may have a more robust appeal. Whether or not the improvement in final colour schemes justifies the loss in flexibility is, as yet, unknown.

The use of linguistic methods for tuning a colour scheme is appealing. It would allow a user to fine-tune a scheme using contextually appropriate language, rather than using the colour space dimensions, although whether this is any more effective (or easier-to-use) than directly manipulating hue, saturation and lightness is another open question.

Colour scheme design using optimisation

The previously described systems based on rules applied them incrementally, either selecting colours or constraining choice as the schemes were developed. However, the bottom-up rule-based approach has inherent limitations. First, starting with the rules and attempting to find colours (usually pairs) that satisfy a required relationship automatically chooses specific colours very early in the design process. The number of possible combinations is huge, and a chosen colour may be later be found to be unsuitable, but in the interim will have acted to constrain other choices. This is a known problem:

> *"expert systems, however, soon get stuck at points where further improvements can be made only by changing many variables simultaneously. These dead ends occur because it is practically impossible to sort out all the effects associated with different multiple changes, let alone to specify the regions of*

> *the design space within which previous experience remains valid."* – *Holland (2009)*

An alternative approach is to defer the evaluation of a colour scheme until it is complete, and if the overall result is not satisfactory, attempt to optimise it. The goal of this optimisation is not to find the global optimum, as there may be more than one, and it is not possible to specify exactly what are the characteristics of the optimal solution. The difficulty in design is centred around humans: *"the most difficult aspects of these design problems are* people*. Designs are usually used by people – we live in architecture, we use and interact with the things around us. We like and dislike things almost at random"* – Bentley and Corne (2002, p36). What is required is a solution that is good enough from the very large number of potential solutions. However, as the search space is so large, it is desirable to use a creative approach that is biased towards finding viable solutions.

Creativity can be defined as *"a mental and social process involving the generation of new ideas or concepts, or new associations of the creative mind between existing ideas or concepts"*[43]. It has long been the popular view that computers can only do what they have been programmed to do, they are predictable and incapable of creating anything new or unexpected. This automatically implies that computational creativity is impossible. If true, this would be unfortunate, as generally, the speed at which computers work greatly exceeds that of humans, and in finding solutions, the larger a solution space that can be explored, the better.

If the creation of novel solutions can be (even partially) automated, it would be possible to find solutions that are unlikely be found by humans. There are problems, colour scheme design among them, where potential solutions are easy to generate, but for which there is no predictable method of finding solutions that are likely to be successful. Humans almost always approach problem-solving with preconceptions of what a potential solution might look like. It is therefore extremely difficult to think sufficiently far "outside the box" to discover totally novel approaches. A computational approach does not suffer from such a myopic view and can find solutions that, from a human perspective, appear bizarre. An illustrative example comes from work by Linden (2002) that designed a radio antenna resembling a crumpled coat-hanger. To a human designer, it doesn't look like an antenna at all, but its performance satisfies the design criteria. The configuration is so irregular that it is extremely unlikely to be tested by a human designer. Another example is the search for new drugs. An automated method of creating and testing compounds (of which there are an extremely large number) can enable the discovery of compounds with desirable properties based on combinations that no-one had thought to test. For an overview of computer-aided molecular design, see Felton (2000).

The approaches used by Linden and Felton are based on evolutionary computation, a method of deriving better solutions from earlier ones by mimicking features of biological evolution. The requirements for creative evolution given by Bentley and Corne (2002)

[43] *Creativity* – from Wikipedia `http://en.wikipedia.org/wiki/Creativity`, accessed June 14, 2009.

include:

a genetic representation: a method of encoding the features of a solution into a sequence of values, such that a particular value (or set of values) determines the expression of a feature in the derived solution,

an evolutionary algorithm: a process that, by modifying a genetic representation, will enable new solutions to be derived,

an embryogeny: a method of mapping from a genotype to a phenotype, from the genetic representation to the corresponding solution,

a phenotype representation: a solution that can be evaluated,

a fitness function: a method of evaluating each solution. This evaluation can be either performed mechanically or by human assessment.

Using these principles, evolutionary systems have been created for a diverse range of application areas, many of which are discussed in Bentley and Corne (2002), which also provides a comprehensive introduction to the various methods of evolutionary computation, including genetic algorithms. Genetic algorithms are widely used and can explore solution spaces, as long as a method of encoding a solution as a genetic representation can be found. The process involves generating many genotypes, each of which represents a different solution, and mapping these to actual solutions that can then be evaluated (Holland, 1975; Marsland, 2009). This requires an objective function, which can be problematic. Some solutions, especially those in design-related fields, are difficult to assess programmatically. Humans may be able to identify solutions, but may be unable to codify the mechanism into an algorithm. An evolutionary approach may still be feasible if a human evaluates the potential solutions. Such a "supervised" approach will be slower than a computational assessment of fitness, but does provide a means of exploring a solution space that might otherwise be impractical. An example of this approach is the two-stage colour scheme design system of Kelly (1996). In the first stage, sets of colour schemes are created without using a colour harmony model, and shown to the user. The user inspects the presented images and tags the "good" schemes. More schemes can then be displayed and tagged. The second stage is similar to the first, but new schemes are based on the more constrained and visually appealing tagged schemes. This enables a slow evolution towards pleasing colour schemes, without requiring a machine-based colour harmony evaluation function. Kagawa et al. (2003) Kim and Cho (2000), in a system for fashion design, also use a person to evaluate the fitness of evolved creations, and Ross et al. (2006) use an evolutionary approach to create textures whose colour distributions are based on sample images.

In contrast to the rule-based approach to colour scheme design, the evolutionary approach does not attempt to codify exactly how a solution is derived, and defers evaluating a solution until it is complete. This helps avoid the very fine-grained, and possibly premature, assessment of features during the design process, such as occurs in the ACE system. With the evolutionary approach, evaluation is deferred until the end, when the solution as a whole is evaluated. This is especially important with colour scheme design, as it is the appeal of the overall scheme that is important, not the appeal of individual pieces.

As is clear from the bottom-up rule-based approach, the exclusion of particular possibilities too early in the development of a scheme can be detrimental, as potentially good part-solutions can be removed. The maintenance of multiple potential candidates within the populations being optimised by an evolutionary approach helps prevent the premature commitment to a particular solution, but does not avoid it entirely.

The unpredictable nature of the interactions between the parts of a colour scheme indicate that the bottom-up rule-based approach, which relies on the codification of these interactions, is unlikely to be successful. The creative or evolutionary approach that uses a "create–evaluate–modify" cycle is more likely to be successful and would allow the schemes, once created, to be optimised. However, for this optimisation to take place without human intervention an algorithmic objective function is required to assess the fitness of an interface colour scheme against the requirements.

2.12 Colour schemes for viewers with colour vision deficiencies

For users with colour vision deficiencies (CVD), some colours can render an interface unusable, and it is therefore appropriate to consider what steps might be taken to ensure that a scheme is suitable for a CVD viewer, or how a scheme can be modified to become so.

As most CVD impairments relate to anomalies in red-green sensitivity, a simplistic solution would be to avoid using the problematic axis in colour schemes. However, as noted in Lumley (2006), this is not an ideal solution. Using only colours from the blue-yellow axis requires the user's display to be able to accurately render variations in blue (all of which will be dark) and yellows (all of which will light). Shortcomings in the user's display could easily result in colours that should be distinct appearing to be the same. Secondly, due to the lack of S-cones in the fovea, and chromatic aberration in the lens, which brings blue light to a focus in front of the retina, focussing on pure blues is difficult. Therefore, relying on blue-yellow sensitivity is not an ideal solution.

Colour and luminance perception are not equally important to vision. Objects usually remain perfectly recognisable if the colour information is removed, as witnessed by the long use of both black-and-white photographs and television. The converse is not true: removing luminance differences and leaving just the colour information can greatly impede recognition. This is illustrated by the pair of images in figures 2.33 and 2.34. Both images have the same colour information, but in the image on the right, the lightness of all colours has been set to the same lightness as mid-grey, without, as far as possible, altering the hue or saturation. Including a monochrome version of the image was considered, but as it appears as a featureless grey rectangle, it was omitted. The various forms of colour vision deficiency affect colour discrimination and matching, but leave the individual capable of detecting differences in luminance (lightness).

Figure 2.33: A full colour image.

Figure 2.34: The same image with all colours forced to the same lightness.

Therefore, as long as the differences in lightness between visually important elements are sufficient to enable them to be distinguished (so that they would remain distinguishable in a black-and-white rendition), no particular correction or adjustment is required for CVD users. If an interface colouring is such that not all interface elements are discernable by lightness contrast alone, the colours can be modified, either when the scheme is being created or when it is being viewed (or both). Several approaches have been used:

render a scheme as it will appear to the CVD user: an example that is intended to be used at design-time is described by Viènot and Brettel (1999). This system is independent of any particular application. It creates replacement colour mapping tables for the video graphics card, effectively converting the complete colour display to produce the image that would be seen by a viewer with a known deficiency. Once designers can see the colour scheme as it will be seen by a CVD viewer, they can modify the colours so that the rendered scheme is acceptable. This approach is also suggested for the colouring of statistical graphics by Lumley (2006). A scheme prepared using this method will be usable by unimpaired users, but unless the designer iterates between using the new mapping tables and not, the colour combinations may appear unusual.

Walraven and Alferdinck (1997) describe an editor to be used at viewing time, for displaying CVD recoloured images. In recognition of the reality that CVD viewers form only a small percentage of the population and that most designers are therefore unlikely to cater directly to their needs, they note: *"we have also developed an expert system that administers an on-screen color vision (self)test, and then uses the test results for the automatic adjustment of the color look-up table to the individual needs of the display user."*

using markup to recolour documents for CVD users: The SmartColor system of Wakita and Shimamura (2005) describes how authors could mark up documents with constraints on colouring. This allows the document to be recoloured for the differing types of CVD in a manner that is consistent with the designer's aesthetic intent and allows for the creation of multiple possible schemes. The criteria proposed are based on defined contrast, colour groups, and distinguishability, along with the ability to exclude items such as photographs from any recolouring. Using simulated annealing (Salamon et al., 2002) and distances measured in the CIELAB colour space, the system creates recoloured images to render semantically important distinctions in informational graphics (e.g. bar-charts). Several colouring are offered, the best and a small number of others, to allow for the case when the designer has not completely defined their intent in the specified constraints.

take an existing scheme and remap the colours for a CVD user: instead of requiring markup within the document, Ichikawa et al. (2004) partition the original image into regions, which are recoloured independently so that the regions are sufficiently distinct, while having minimal impact on pictures of natural scenes. A system oriented more towards documents is that of Jefferson and Harvey (2006). They describe how a document may be recoloured so that the perceptual distances in the original document are preserved, but the recoloured version is suitable for a viewer with a known colour vision deficiency. Ro and Yang (2004) describe a system for the 75% of CVD users who are anomalous trichromats, which will preprocess the colours by applying an inverse transform, so that the colours perceived by the user after having been incorrectly "processed" by their visual system, appear as intended.

Of the schemes that are not designed with CVD viewers in mind, some will be acceptable, and some become so if the colours are adjusted. The adjustments could use (if available) colour scheme-specific markup embedded by the designer to provide a recolouring suited the viewer. However, due to the upfront effort required on the part of designers, the wide variety of tools used to create interfaces, and the relatively small percentage of viewers affected, the widespread inclusion of embedded CVD-specific information is unlikely.

During colour scheme design, the only generally applicable guideline to aid CVD viewers is that of maximising lightness contrast. For schemes not designed with them in mind, CVD viewers must either put up with some suboptimal images or use software that remaps the colours to suit their impairment.

2.13 Summary

The colour scheme of a user interface can significantly affect a user's perception of a software artifact, the credibility of an online site, and the user's desire to use (or continue to use) a system. It is therefore important to ensure that an interface colour scheme gives a positive impression. However, this is complicated by the difficulty in defining, a-priori, exactly which features are important. It is difficult to define what makes particular interface colour scheme pleasing or appropriate, suitable for a particular context or culture, or pleasing to a particular individual.

Colour space is three dimensional, and the representation of colours may be encoded in many ways, some mathematically tractable, some not, and of these colour spaces, some are more useful in the design of colour schemes than others. Human colour vision is quite tolerant of changes in the viewing conditions. This, combined with the standardisation of display technology, has meant that electronically encoded images can be decoded and displayed with widely acceptable colour rendering, with no particular effort on the part of the viewer. This is not so for those with colour vision deficiencies, but most impairments cause altered colour perception, not colour-blindness, and affect only a small percentage of the population.

The light from an object might be thought sufficient to define its colour, but colour itself is a perceptual phenomenon, and this perception is affected by factors other than the spectral power distribution of light from an object. These factors include the saturation and adjacency of other coloured objects and the viewing conditions. These factors, along with the "composition" of a scene, affect the holistic impression of a set of coloured objects when seen together. If the effect of the combination is pleasing, the arrangement is said to be harmonious.

Colour harmony is therefore a concept that is easy to understand intuitively, but difficult to define more precisely. Nevertheless, the factors that appear necessary for an arrangement to appear harmonious have been studied and, while the findings are not in complete agreement, there are models that, given set of objects, will determine the appropriate values of lightness and saturation for each object. Hue can then be chosen to give a final colour scheme with broad appeal.

Methods of selecting harmonious colour arrangements have been derived by considering the aesthetic aspects of images, but these methods do not include factors necessary to ensure that an interface will be usable. The most important omitted consideration is the readability of text. This has been found to be strongly dependent on the contrast of the text with its background. Simple colour space distance is insufficient to guarantee readability, as lightness contrast has been found to be more important than contrasts of hue or saturation. Importantly, there are contrast guidelines that can be used to ensure text is readable.

The methods used to create interface colour schemes range from the completely manual selection of colours for individual interface elements, through to completely automatic systems that produce a single final colour scheme. Manually selecting individual colours is the most flexible, but is also the most problematic for artistically naïve developers. Completely automatic systems also have problems: their schemes are not always successful, and they do not allow the sometimes subtle colour preferences of the developer to be incorporated in the colouring. To aid artistically unskilled developers, the optimum method would be somewhere in between: it would guide the user towards schemes that are likely to be appealing, and it would have enough flexibility to allow the developer's judgement to influence the colours selected.

The systems for automated colour scheme creation discussed in this chapter use several different methods. Some incorporate sets of rules relating to colour harmony, which are by their nature generic, and therefore incomplete. Some allow the specification of constraints pertaining to the interface itself, and some constrain the colours to suit the intended audience. Some have methods for searching or constraint satisfaction, performed either automatically or manually. None of the systems intended to produce a single complete colour scheme can be relied upon to be successful, due to the incompleteness of the colour harmony heuristics, unstated requirements relating to the intended use, and the subjectivity of human colour preference. Rather, as explored in Terry and Mynatt (2002), it is highly desirable to combine the guided creation of new designs with iterative assessment and the ability to explore multiple alternatives, a facility available in few of the surveyed systems.

It is now time to investigate, in more detail, how a system could be structured that allows such a generative approach to colour scheme creation. It would blend automatic design with human exploration and assessment to aid in the design of user interface colour schemes. The next chapter will describe how the building blocks of such a system may be designed, and show how these could be composed into an architecture to enable the design of user interface colour schemes.

Chapter 3

The conceptual foundations of an automatic interface colouring system

This research explores the extent to which the design of user interface colour schemes can be automated to aid software and web developers who are unskilled in the use of colour. The aim is to help choose a set of colours for the individual elements within a user interface so that the resulting scheme will please the developer, suit the intended use, and appear to be well designed. However, the colours within a scheme, while defined for individual interface elements, are best not chosen in isolation, as the effect of a colour scheme is holistic. Therefore, the method used to select colours for each interface element should take this holistic impression into account.

Artistic models of colour harmony attempt to model this holistic impression, or more accurately, to define relationships between the colours of elements in a scheme so that the holistic effect is pleasing. However, the models do not include interface-specific considerations, nor the suitability of any particular hues, which are important if the interface is to be both usable and be coloured appropriately for its intended use. Therefore, any attempt to fully automate the design of an interface colour scheme using such models is unlikely to be successful. The difficulty in automating the design of a colour scheme becomes clear when the number of interacting factors affecting the choice of colours in an interface is considered. These include:

factors relating to the structure of a colour scheme:

- the colours of individual objects are usually chosen so that strongly saturated colours are not assigned to large areas to avoid garish schemes that can be visually overpowering, distracting, and unpleasant to use.
- when seen together as a complete scheme, the holistic effect of the coloured elements should have a pleasing aesthetic effect.
- the colours that may be used for elements are mutually constrained. The flexibility in colour choice will decrease as objects are coloured, as the colours

of previously chosen elements will limit the aesthetically appropriate colours that may be used for the remaining elements.

interface-specific factors:

- non-decorative text should be readable.
- the colouring of items should reflect any structural and semantic relationships in the user interface.

"soft" factors:
The appropriate colours for a particular interface may be influenced by ill-defined factors, such as:

- the need to elicit a particular impression from the viewer in response to the overall scheme (e.g. restful, fresh, sophisticated).
- the prior associations given to particular colours.
- the emotive impact of particular colours.
- the culture of the intended audience.
- subtle aesthetic effects. As a simple example, a scheme may look better when colours are used in a particular way; e.g. using colour c_1 for element$_1$ and c_2 for element$_2$ may look better than the other way around, even if the areas of element$_1$ and element$_2$ are the same and the overall balance is unaffected.
- the subtle effects that can cause some hues to appear more suitable or pleasing than others (e.g. even among all the greens that may appear "fresh," some will appeal more than others).

The first two items (colour scheme structure and interface-specific factors) pertain to the relationships between the colours of the interface rather than the particular colours used. It is these colour relationships that the artistic colour harmony models partially address. The "soft" factors relating to the actual hues used in the scheme are not addressed by the artistic colour harmony models, but are important for an appropriate interface colouring, and therefore a means of allowing for such effects should be included. This can be addressed by designing a colour scheme in stages.

The first stage would consider the only lightness and saturation relationships between the coloured items in the scheme, but not hue. This would result in a colour scheme devoid of actual colours – an abstract colour scheme – that satisfies the inter-element relationships for colour harmony and interface-specific factors.

Once these relationships have been determined, the second stage could address the assignment of appropriate overall hues to satisfy the soft factors. However as the selecting the hues should not alter any of the inter-element relationships defined in the first stage, the colours in a scheme will change in concert – setting or changing the colour of any one element will change them all – therefore, the adjustment of hues in the second stage changes the colour scheme holistically.

The combination of a harmonious inter-element relationships and holistic colour selection would result in an overall colour scheme that satisfies both the requirements of a usable and harmonious user interface colour scheme and the "soft" requirements related to the intended audience and aesthetics. However, the selection of hues that satisfy all the subtle and interacting constraints is difficult to define algorithmically, but is something that human can discern with little effort by inspection. Therefore, the proposed approach is for the automatic creation to pertain to the structure of the colour scheme – the relationships between the coloured elements – and the selection of actual hues to be left for the developer, who will automatically avoid inappropriate choices.

Developers, even those with limited colour expertise, are likely to have preferences about how a scheme should be coloured, both in terms of the relational aspects (the first stage), and the appropriate hues (the second stage). They should therefore be able to specify such preferences, without needing or using any specialist knowledge. Minimising the input required from developers and extracting as much data as possible from the interface to be coloured will make it less likely that developers may misinterpret, or being unable to supply, any details relating to the colour scheme.

This chapter explores these ideas and partitions the functionality into conceptual building blocks. The next chapter will show how these conceptual blocks can be mapped to an implementation which can be used to generate schemes to test the viability of the approach.

Although the ideas explored in the thesis are related to aesthetics, it is recognised that there other important contributions to the aesthetic quality of an interface, such as the layout and placement of interface elements and the choice and size of fonts. Such considerations are outside the scope of this project, and are assumed to have already been optimised.

3.1 Augmenting colour harmony heuristics for use in GUI design

Conventional colour harmony heuristics, such as the use of sets of colours related by a perceptual ordering, or the requirement for visual balance, are based solely on aesthetics. They do not address the semantic relationships between the elements in a user interface and they may therefore select colours that harmonise, but impair the usability of the interface. In a colour scheme that has been produced by following such heuristics, there is no guarantee that controls will be visible or that textual elements will be readable against their backgrounds. In an interface, usability considerations – the readability of the text and the distinguishability of controls – are paramount. Therefore, the existing theories of colour harmony must be extended to guarantee that such semantically important detail is clearly visible.

Semantic and aesthetic considerations

The starting point for the colouring process is an interface that is complete, apart from its colour scheme. The aesthetic appeal of a colour scheme is affected not only by the colours of the objects themselves, but also by how those objects are arranged with respect to one another. However, even though the object placement is fixed, differing colours will produce schemes that are more (or less) pleasing.

The visual impact of the placement of coloured objects falls into the realm of art and aesthetics. This topic has been explored by many artists. For example, Mondrian created many works exploring the visual effect of the placement of solidly coloured areas, samples of which are illustrated in workbook of colour exercises by Hornung (2005). Kandinsky also explored colour and form in his later work (Beckes-Malorny, 2007). The research into evaluating aesthetics is ongoing (Ngo et al., 2002), but as the placement of the interface elements is predetermined, but the measures of aesthetics not related to colour harmony are not relevant here. The only degree of freedom is colour.

Interlinked with the effects of object placement are the effects of simultaneous contrast, where the apparent colour of an object is affected by the colours of nearby objects (see sec. 2.2.6 for details). There is no well-accepted algorithmic method for the evaluation of simultaneous contrast effects, nor is it clear that these effects need to be included in a system for user interface colour scheme design. Therefore, the proposed architecture does not attempt to incorporate the visual effects of object placement (including simultaneous contrast). Rather, it is left to the human developer to select colour schemes from those created automatically where the visual arrangement is appealing, and ignore those schemes where the colours of nearby and juxtaposed elements do not enhance the overall scheme.

The placement of controls in an interface is not arbitrary: the designer of the interface will have placed the controls to visually encode relationships between form and function. Two techniques are commonly used: (i) placing related items in visually distinct groups, either by proximity or by using container controls with borders; (ii) using the same or similar colours for related controls. The developer must be able to ensure that these semantically important groupings are reflected in any subsequently generated colour scheme.

Incorporating developer-specified aesthetic constraints

How, then, should these underlying semantics be expressed? Program development and web design environments are not designed to capture information oriented towards colour scheme design (e.g. the desired emotive impact of the interface or details of the intended user group). While it would be ideal to capture this information automatically, it is not easy even to define what information is actually required.

As discussed in section 2.11, the visual and emotive appeal of colours are not only difficult to define, but their fuzzy nature would also make them difficult to encode in any comprehensive way, and any simple rule-based approach would be prone to error. This difficulty in definition and encoding also applies to the suitability of a colour scheme for a particular audience[1]. These ill-defined factors are ones that a human can assess quickly, without necessarily being able to explain how.

To allow the automatic creation of colour schemes appropriate for a particular audience, a significant amount of ill-specified data (e.g. the colours to be included[2], cultural context, age range of the target audience, gender, emotive associations, real-world associations etc.) would be required from the developer before scheme creation could commence. This is inconsistent with the requirement that a minimal amount of non-specialist input be required from the developer, so this requirement is addressed in two ways; requiring some input from the developer (the next section) and allowing any automatically created scheme to be adjusted after creation (sec. 3.9).

3.2 The developer's input

Two remaining types of information are needed from the developer. The first is a way to indicate how controls are grouped (usually indicating that their functionality is related), so the colours in the created schemes can reflect this grouping. The second is a way for developers to define constraints to ensure that the colouring of the interface elements does not detrimentally affect usability.

There are interface elements such as user controls, where colour distinction is important for semantic rather than aesthetic purposes. To ensure that an interface colour scheme is usable, there should be a mechanism for the developer to define any pairs of interface elements that should be coloured so that they are mutually distinguishable. This colouring constraint could be used to ensure the colour difference between the specified pairs of interface elements is large enough that there is no possibility of the objects being hard to find (or invisible) due to insufficient foreground–background contrast. The developer may also choose to use this constraint for aesthetic purposes, to force particular interface items to be coloured differently.

The user-defined constraints – the decision about whether or not to colour some items identically or to force others to be differently coloured – are value judgements based on aesthetics, interface semantics, or both. It is common for multiple elements in a user interface to use the same colour, so there should be a mechanism to allow the designer to require this, and also to (optionally) express preferences on the relationships between the colours of particular controls (or groups of controls). These groupings and the need for distinctions between elements are a natural part of interface design, and specifying them would not require any colour design expertise. Interestingly, for many interfaces, usability will not be greatly affected if these constraints are not specified at all. However, their inclusion increases the likelihood that the generated overall colour schemes will be pleasing. When defining the groupings and constraints, it is likely that several iterations would be required, with groupings being altered and constraints

[1] This could be a "user group" (for application software), or "viewers" (for web sites), but "audience" seems to encapsulate both meanings and is commonly used for the recipients of a creative work.

[2] There are exceptions. Some colours (e.g. *Kodak*™ yellow) are very precisely defined, but more usually a textual description of a colour (e.g. "leaf green") can encompass a very large number of actual colours.

added or deleted until the final scheme meets with the developer's approval. This is no reflection on either the developer or the method: iteration is normal when finalising a creative design.

If the system works as intended, even without user-defined constraints and groupings, all the schemes produced would conform to colour harmony heuristics and readability constraints, and therefore should be usable and harmonious. However, without the addition of user-defined constraints and groupings, the colourings would not reflect the semantics of the interface and would be open to obvious improvement, leading the developer naturally towards refining the constraints and groupings to improve the created schemes. Moreover, if the redefinition of groups and constraints was interactive, and the subsequent creation of updated colour schemes was sufficiently rapid, this would promote learning through exploration. It also mirrors the iteration inherent in graphic design.

3.3 The information resulting from the interface characterisation

Following the user-defined grouping of the interface items and specification of any required distinguishability constraints, the following information could be derived to allow the creation of the interface colour schemes:

- a list of colourable items. A colourable item is either a single interface item or a group in which all interface elements in the group[3] are assigned the same colour - a "same colour group". These groups are meta-elements with a single colour and a composite area. In effect, they are no different from a single larger interface element.
- a list of interface item pairs that the developer has specified should be distinguishable by colour alone.
- a second list of interface item pairs (textual interface elements and their backgrounds) whose colours should be coordinated to ensure easy readability of the text.

These would be sufficient to ensure that groups of interface elements could be coloured as a set to reflect the interface semantics, and that elements (and groups) that should distinguishable and readable could be coloured accordingly.

3.4 The concept of an abstract colour molecule

As discussed in section 2.9, there have been many heuristics for harmonious colour schemes. The concept of using an ordered progression of colours appears important, and these sets may be derived by choosing intervals between two (sometimes more)

[3] The members of a group are usually real interface elements, but there is no reason why groups cannot be nested.

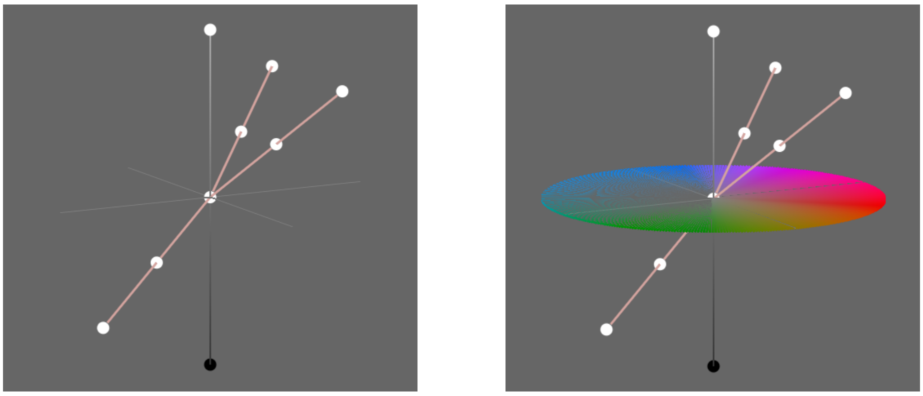


Figure 3.1: A wireframe tilted at 45° to the vertical lightness axis (white at the top, black at the bottom) will allow the colours along the wireframe path to include colours with varying lightness (from the different vertical position) and, as can be seen from the figure on the right, varying saturations from the differing radial distances.

extremes in the colour space. In general, the number of intermediate colours that exist between the extremes greatly exceeds the number of items to be coloured, so an automatic colour selection system has a good deal of flexibility in choosing amongst them.

Two visualisations for colour schemes based on interpolated positions on a wireframe have been found to be intuitive and helpful. When discussing the choice of colours, it is convenient to think of the progression of colours as a notional structure of wires stretched between the extreme colours in the colour space, and the colour of each interface item as a bead that can slide along the wire(s). Once the positions on the wires have been chosen, it is convenient to think of the colours as atoms at fixed positions in a rigid molecule. The geometry of the molecule is (now) fixed, so the relative positions of the "atoms" are not affected by a changes in the molecule's position or orientation.

To be of practical use in the selection of harmonious colours, these models need to be capable of representing variations in hue, saturation and lightness, and the colour space in which they are deployed needs to be perceptually uniform. The first of these requirements implies a three-dimensional colour space; colour wheels that represent hue variation alone, and colour discs that represent variations in hue and saturation are not sufficient. If the colour space also is perceptually uniform, then a given Euclidean distance in the space will represent the same amount of perceived colour difference in any region of, and at any orientation within, the colour space. This property is one of the foundations of the approach to colour harmonisation explored in this research project.

The most widely-used colour schemes are the monochromatic, complementary and split-complementary schemes, all of which can either be represented as or approximated by lines or line-segments in colour space (i.e. wireframes). Colour schemes derived from curve-based paths are also possible. Allowing the developer to choose one of these colour schemes as the basis for the colours used in the interface will allow the selection of the related wireframe, which can be specified in geometric terms.

The orientation of the wireframe will determine the hues and also the variations of lightness and saturation that are possible. For example, if a complementary or split-complementary wireframe was completely horizontal within the colour space[4], then all points on the wireframe would have the same lightness, but varying degrees of saturation. If the major axis of the wireframe was tilted, for example, at 45° to the horizontal (as shown in figure 3.1), then positions along the wireframe would correspond to colours with different degrees of lightness and saturation.

Rotating a colour molecule in a perceptually uniform colour space does not alter the perceptual differences between pairs of colour atoms or the harmony of the overall colour scheme represented by the molecule. Therefore, defining the arrangement of atoms within a rigid colour molecule is equivalent to defining an abstract colour scheme. Defining the position and orientation of this abstract scheme fixes the position of its atoms in colour space, which maps each atom position to an actual colour, and so defines a real colour scheme. The separation between abstract and real colour schemes, and the invariance of colour differences under rotation are intended to produce two effects. First, they should make it possible to design harmonious and usable colour schemes without fixing on a set of real colours, and secondly, they should they make it possible to "tweak" a colour scheme without changing its degree of colour harmony, by altering the position and orientation of the colour molecule.

The use of interpolation to produce ordered colour sets

A range of colours can be selected by points along the wireframe path: when the wireframe is placed at a particular location and at a particular orientation in colour space, each point will correspond to a unique colour. This satisfies two of the criteria that have be found helpful in selecting harmonious colours: the use of the artistic community's harmonious wireframe shapes and the inclusion of visible order into the selected colours. Colours selected from points on the wireframe will form a palette that can be used as the basis for a harmonious colour scheme.

Three dimensional colour space visualisations typically show the external surface, which is good for showing the more saturated colours but not the desaturated colours that are closer to the centre. A visualisation of the double cone HSL space with geometry adjusted to allow internal colours to be seen was published in Varley (1980). This graphic was used as the basis of a dynamically adjustable software implementation (shown in fig. 2.27). This inspired a generalisation (shown in figure 3.2) in which a spherical cutaway with a resizable wireframe is placed on a tiltable plane, with the colours selected from the bead positions. This colour selector – the "Chromotome" – is based on research by the Massey University HCI Group and is discussed in more detail in Moretti et al. (2004). Changing the tilt angle of the wireframe (either sideways or front-to-back) or the size of the wireframe will alter the six[5] colours selected by the beads on the wireframe, without altering their relationship in colour space.

[4] assuming that the lightness axis is vertical

[5] Only the beads on and at the end of the wireframe segments are used to select colours. The front- and right-most beads can be used as controls to alter the tilt angles and size of the wireframe plane.

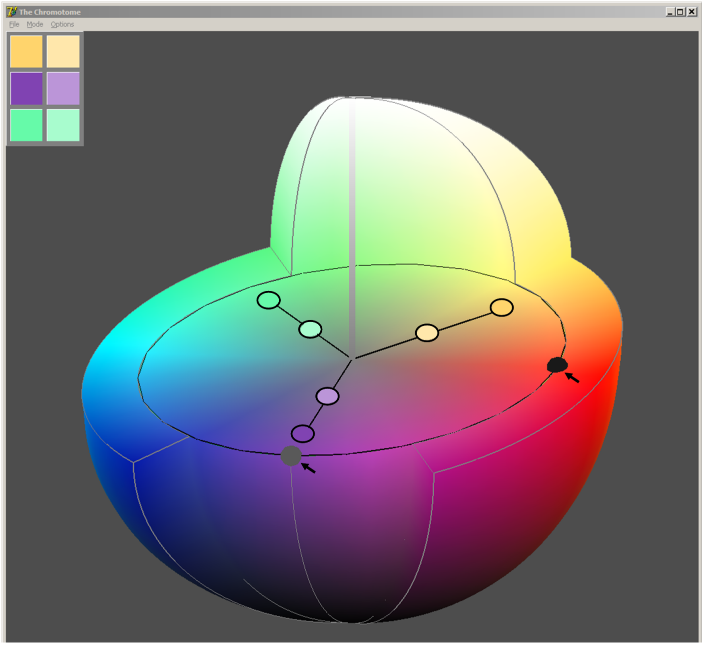


Figure 3.2: The Chromotome: a cutaway view of the HSL colour space showing the internal colours, illustrating how the placement of a wireframe with the dots (beads) on the wireframe indicating positions in the colour space, which correspond to sets of colours. The wireframe lies on a plane that may be tilted to vary the lightness of some beads, and the size of the wireframe itself (indicated by the circle on the plane) may be altered to vary saturation. The six circles outlining the beads and the two arrows indicating control handles have been added to the image for clarity in this thesis. Software by the author.

Designing colour schemes in the abstract, without considering colour

The beads on a wireframe denote coordinates in colour space. As the wireframe is moved, the colours denoted by the bead positions will change. The idea of a movable wireframe shape that can be arbitrarily positioned to create colour schemes presupposes that the relationships between colours at fixed points on the wireframe are invariant under rotation, and that all wireframe rotations yield equally acceptable colour schemes.

These ideas, if correct, are very useful in a system for colour scheme design. They allow the structure of the colour scheme (the relationships) to be considered and manipulated without knowledge of the actual colours. The non-colour-specific scheme is designed to satisfy structural constraints, and the resulting set of coordinates can be mapped to many differing sets of real colours. The actual colours have been abstracted out of the design of a colour scheme.

There is evidence that both suppositions (invariance and acceptability under rotation) are true: fixing the relationships between the colours under rotation can be achieved by performing the rotation in a perceptually uniform colour space; and the independence of colour balance from hue is supported by the experimental results of Guan (2002). Choosing different positions for the items along the wireframe will allow a variety of palettes to be created.

This approach to colour harmony does not, as yet, incorporate any way of allowing the properties of interface items to influence the intensity of colours. Large areas could be allocated highly saturated colours, which would be unpleasant and tiring to look at – undesirable attributes in an interface colour scheme. To avoid this, the method used to select positions on the wireframe would need to take into account the area of the interface elements. Therefore, it is not possible to design a harmonious colour scheme (even an abstract scheme) without knowledge of the actual interface the scheme is intended to colour.

The wireframe path constrains the set of colours that may be used. From this, particular subsets may be chosen that satisfy a colour balancing algorithm that, with a wireframe, encode the colour harmony heuristics. The working assumption[6] is that an algorithm that uses area to derive appropriate values of saturation and lightness, which when combined with wireframes, can be generalised to the more complex environment of user interfaces and can be used to create harmonious user interface colour schemes.

Therefore, once the developer has selected a wireframe and the areas of the various interface elements have been determined, there is sufficient information to start the creation of a colour scheme. An outline of the data used and created during interface characterisation is shown in figure 3.3.

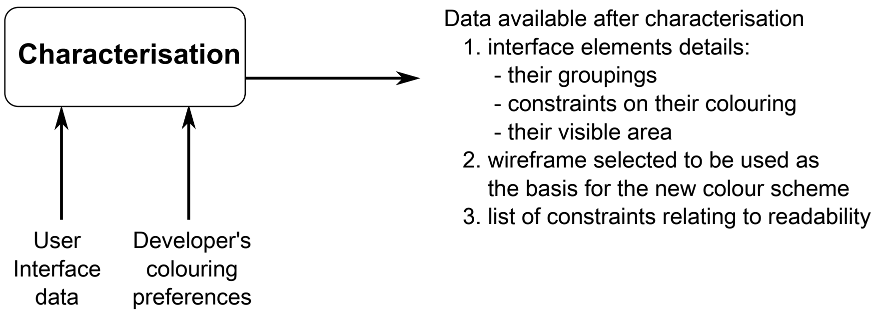


Figure 3.3: Data from the user interface is extracted, and used to allow the developer to specify groupings and constraints between the colourings of the interface elements.

3.5 The conflicting demands of colour harmony and GUI design

For an interface to be usable, some pairs of elements must be mutually distinguishable, and text must be readable. There are a very large number of possible colourings of an interface, and some of these will be closer to satisfying both the colour harmony and the user-defined constraints than others. User-interface-specific concerns, such as readability and distinguishability, are not addressed by any aesthetically-derived theory of colour harmony, and therefore, existing theories of colour harmony must be extended to incorporate them. However, adding user-defined aesthetic constraints[7]

[6] which will be tested experimentally and the results presented in chapter 6.
[7] here, meaning grouping and distinguishability.

and automatically derived readability constraints alters the design of an interface colour scheme from a problem of balancing areas of colours chosen from a wireframe (a problem with many exact solutions), to one with no solution, as it may not be possible to satisfy the requirements of colour strength balance, distinguishability and readability at the same time. The creation of a GUI colour scheme requires the satisfaction of multiple competing constraints.

The presence of an underlying wireframe (e.g. complementary split-complementary etc.) is implicit in many of the models of colour harmony (sec. 2.9). However, there are more constraints when selecting colours for use in an interface colour scheme than when selecting colours for purely artistic purposes. Therefore, it may be necessary to select colours off the wireframe to ensure sufficient colour separation to satisfy the readability and distinguishability constraints. If this should be necessary, for aesthetic reasons, it is desirable to choose colours that are as close to the wireframe as possible. This constraint can be named *wireframe alignment* – the need to constrain the colours used in the scheme to those colours defined by the wireframe, or as close to it as possible.

There are now four requirements, or equally, constraints. The first two are aesthetic: (1) colour strength balance and (2) wireframe alignment; and latter two are pragmatic: (3) readability and (4) distinguishability between pairs of colourable items that may be defined (if desired) by the developer. This naturally leads to the question of the relative importance of the four factors to the final colour scheme. If they are not equally important, what weight should be given to each? The importance of each factor to the appearance of a colour scheme requires a perceptual evaluation, and so the weights would need to be assessed from the results of experimental trials.

3.6 Colour schemes are independent of colour

Fixing the positions of the colourable items in colour space does two things. Firstly, each position in colour space corresponds to a unique colour, so fixing the positions of interface elements in the colour space defines their colours. Secondly, it enables the evaluation of the relationship between any pair of items, such as their separation in colour space or their difference in lightness. These values enable an assessment to be made of how well a particular set of colour space positions satisfies the required constraints.

The "bead on a wireframe" is a useful metaphor when thinking of moving items on a wireframe, but once the positions are chosen, another visualisation seems more appropriate: the image of an arrangement of fixed items on a wireframe is not visually dissimilar to the ball-and-stick models used by chemists to visualise the arrangement of the atoms within a molecule. To simplify the discussion, it is convenient to introduce two terms:

a colour atom: a colour atom associates a colourable item (i.e. a single interface element or a same-coloured group) with a position in colour space.

The properties of a colour atom are its coordinates in a colour space, the item (or group of items) that it refers to, and the area of the related interface element (or for groups, elements).

The atoms introduce a level of abstraction: the properties of a colour atom are its *area*, its *position in colour space* and whether it (or for groups, any of its members) is *textual*. A colour atom may represent one or many interface items but, when evaluating colour harmony relationships, it is not necessary to know which. The coordinates may refer to a real or an abstract colour space, although if the space is abstract, a mapping from the abstract coordinates to real colours is necessary before the related interface element can be recoloured.

Moving a colour atom in colour space alters the colours of the associated interface element; if the element is a same-coloured group, the colour change is propagated to all interface elements associated with the group. If a mapping from abstract to real colour space coordinates has been defined and the interface is available, any change in the position of a colour atom can be used to update the displayed colour scheme.

a colour molecule: a colour molecule is a set of colour atoms. As each colour atom has a position in colour space, the molecule defines a set of atoms and implicitly, the spatial relationships between them. The set of atom positions therefore also defines a set of colours (a palette). As each colour has an associated object, the colour molecule therefore represents a colour scheme. This scheme can be translated and rotated within the space, without affecting the colour relationships between the atoms (and thus, between the colours of their corresponding interface elements).

Using a colour molecule to represent a colour scheme provides a useful level of abstraction. A colour molecule defines the geometric relationships between the colours of the interface elements in colour space but, since it could be positioned and oriented anywhere in the colour space, those relationships are independent of actual colours. The colour molecule is thus a type of generalised colour scheme.

Since the four colour constraints can also be expressed as geometric relationships within the colour space, a colour molecule can be used as an abstract colour scheme; its geometrical properties can be plugged into an evaluation function to evaluate the abstract colour scheme. The result of this evaluation will be independent of the molecule's position and orientation within the space. Ideally at least, the quality of real colour schemes should not be affected by the repositioning or reorienting of the molecule in the space; although the colours will change, the relationships between them will not.

The evaluation of how well a colour scheme satisfies all four constraints can only be arrived at by considering a complete colour molecule, not the colour atoms individually. While wireframe alignment (a measure of how close a colour atom is to the wireframe) can be evaluated for individual atoms, the readability and distinguishability constraints relate to the distances between pairs of colour atoms, and the assessment of colour balance must consider all the colour atoms within a molecule. Combining these factors to give a single score enables an overall assessment of how well a particular colour scheme, as represented by a colour molecule, satisfies all four requirements.

The data from the interface characterisation phase is sufficient to create colour molecules in an abstract space (figure 3.4) and now that the evaluation of how well a molecular arrangement satisfies the constraints has been outlined, methods for positioning the atoms within a colour molecule can be considered.

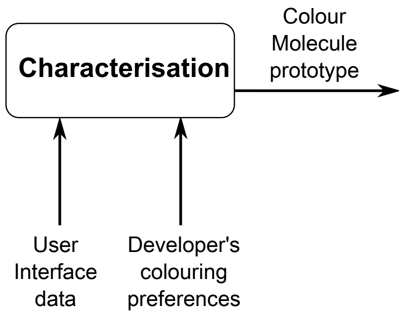


Figure 3.4: The data from the interface characterisation is sufficient to allow the construction of colour molecules (each representing a colour scheme) in an abstract colour space.

3.6.1 Initial colour atom positions within a molecule

Many different colour schemes could be created simply by dividing the wireframe path into equal intervals and randomly assigning atoms to the end-points of the intervals. This would maximise the colour space distance between adjacent colour atoms, and would therefore also maximise distinguishability, and possibly readability.

There are as many differing colour schemes as there are orderings of atoms on the wireframe (e.g. $n!$ for n colour atoms on a single line wireframe). Some of these orderings will correspond to colourings that are closer to satisfying the constraints than others. In some, a textual item and its background will be sufficiently close to each other in the colour space to render the text unreadable, in others the readability may be better, but the colour strength balance may be worse[8].

3.7 Improving a colour scheme – optimisation

In general, it is impractical to evaluate all possible orderings of equally spaced atoms. Even if it were possible, it might be futile, as it may be impossible to satisfy all of the constraints without moving the atoms from the equally spaced positions.

An alternative approach is to try and find positions for the atoms on the wireframe that satisfy the constraints. There is no guarantee that this is possible, but as there are many more combinations of inter-atomic distance[9], it is more likely that a solution

[8] For example, a colour scheme with many items of text on one side of a complementary scheme with a much larger background area on the other side and at the far end of the wireframe. This could have a good readability (from the large vertical (lightness) difference), but the colour strength balance would be poor.

[9] All constraints are expressed in terms of colour space distances.

could be found. To test for this possibility, an optimisation phase could be included to reposition the atoms to maximise the degree of fit for all four constraints simultaneously.

There is a very large number of possible colour schemes and the schemes found using this approach may not be globally optimal (for any given set of criteria), but the aim is not to find some theoretically optimal "best" scheme, it is simply to find a set of colour schemes that are good enough. It would be sufficient to find colour schemes that satisfy all four constraints. Such a scheme would satisfy the developer-defined constraints and the colour harmony heuristics. It would therefore be a "good" colour scheme from the point of view of conventional colour harmony theory, and would also be readable and have distinct colourings, so it would be usable as an interface colour scheme.

Any scheme resulting from an optimisation incorporates only the most basic of the developer's aesthetic preferences. However, there is no difficulty in producing a number of alternative solutions; this would make it possible for the developer to choose a personal favourite from a set of usable and harmonious colour schemes. It is therefore desirable for the optimisation to return multiple solutions. This could be done either sequentially (if one solution isn't appealing, present another), or in parallel, by presenting a small number of optimised schemes at the same time.

3.8 Deriving real colours from abstract colour space coordinates

The utility of an optimisation process is contingent on the existence of a mapping between colour coordinates in the abstract colour space and colour coordinates in a real colour space. For this to be useful, it is necessary to know that abstract colour space coordinates can be mapped to real colours and what transform will be used, but it is not necessary to know which actual colour will be assigned to any particular colour atom. However, it is necessary to define, prior to optimisation, which abstract colour space axis will map to the real lightness axis, so that it is possible to determine whether a particular arrangement of colour atoms satisfies any readability constraints, as readability is largely determined by lightness contrast.

Therefore, the data from the interface characterisation phase, when combined with a colour harmony model and optimisation (to ensure the constraints are met), are sufficient to allow the creation of colour schemes (represented as colour molecules) in an abstract colour space (figure 3.5).

There is nothing in the optimisation that requires a particular hue to be associated with one end of the wireframe, so the hue of any single colour atom[10] can be freely assigned, by rotating the wireframe[11] which will create a new set of equally harmonious colours at each rotational increment.

If rotating the colour atom coordinates does not affect the apparent visual differ-

[10] excluding those on the lightness axis.

[11] "rotating the wireframe/molecule" is short for "rotating the colour atoms while maintaining their relative positions" (as though they were attached to a wireframe).

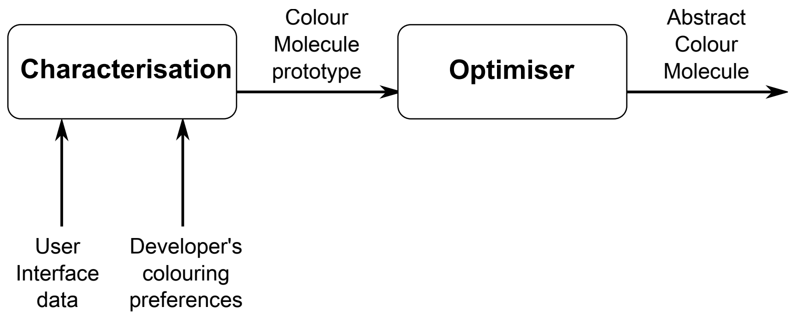


Figure 3.5: The initial colour schemes are differing orderings of fixed wireframe spacings that can be considered as candidates or prototype colour molecules. These are optimised to create abstract colour molecules representing colour schemes that are harmonious, readable, and incorporate the design characteristics and constraints specified by the developer.

ences between the atom colours, then the colours used in the scheme may be changed at will and, importantly, without requiring the optimisation to be repeated. However, the colour relationships embodied in the abstract colour atom coordinates will only be preserved if the space in which this rotation is performed is perceptually uniform.

Device-specific colour spaces such as sRGB (and its transforms such as HSL and HSV) are not perceptually uniform. While any abstract colour scheme must eventually be mapped to an RGB space for display, the colour space within which the rotational transforms are performed cannot be a perceptually non-uniform space, such as sRGB. The CIELAB colour space could used instead. It is mathematically tractable, is close to perceptually uniform, and has transforms to and from RGB. The colour atom coordinates could therefore mapped from abstract colour space to CIELAB, and from CIELAB to an RGB space for display.

A reason for not using CIELAB as the colour space in which the colour molecules are created and optimised would be conceptual simplicity. The CIELAB space (or any other perceptually uniform space) has a very odd shape, and designing colour schemes in such a space would force the actual colours used in a scheme to be considered as part of the design. This is at odds with one of the research goals of this work – the design of complete and freely recolourable schemes. It would be simpler if the schemes were designed in an ideal abstract space, and pragmatic considerations, like mapping these schemes to real colours, handled at a later stage. The implications of this decision are detailed in section 4.7.

3.9 Personalising the schemes created by the optimiser

It is possible that the optimised abstract colour atom coordinates, when mapped to real colours, result in hues that do not appeal to the developer, or are culturally or stylistically inappropriate for the intended audience. It is therefore necessary to allow the developer to personalise the colour schemes produced by the optimiser. The primary form of adjustment would be to tailor the hues (as there are subtle effects that the broad model of colour harmony encoded in the fitness function does not attempt to address).

The hues of elements within the scheme and their saturation are two obvious colour parameters that strongly affect the visual appeal of a colour scheme. A mechanism to adjust these to suit the developer and the target audience would introduce a level of personalisation that would cater for personal preference without affecting the overall colour harmony designed into the scheme during the optimisation.

The personalisation of a usable and harmonious colour scheme is a desirable final phase. It would allow developers to tweak[12] the colour scheme to their own taste. The inputs to the personalisation would be the set of abstract schemes output by the optimiser that satisfy the requirements of colour harmony, wireframe alignment, readability and the user-defined aesthetic constraints, and the outputs would be real colour schemes of defined hue and saturation.

3.10 The modular structure of the proposed architecture

The major functional blocks of the proposed architecture and the dataflow between them are illustrated in figure 3.6. The elements within the figure can be grouped by block (within the diagram) into two stages:

creation of abstract interface colour schemes:
This involves

- *characterisation:* data from the user interface components (1 in fig. 3.6) and the user's specification of their preferred groupings and which items should be differently coloured (2) are combined to create a variety of potential colour schemes (3).
- *optimisation:* potential colour schemes (3) have their items repositioned in an abstract colour space to find arrangements that allow the colour balance, wireframe alignment, readability and distinguishability constraints to be satisfied simultaneously, giving optimised abstract colour molecules (4).

adjustment of an optimised scheme:
One of the abstract colour molecules (4) is selected and the colour scheme is holistically adjusted (without requiring reoptimisation). This involves

- *rendering the scheme:* the values of the colour personalisation controls are used to map an abstract colour scheme to sRGB colours (5) and these colours are used to update the colours within the displayed interface.
- *evaluating the scheme:* the user evaluates the displayed scheme (6) and considers whether it uses colours that appeal and are appropriate for its intended use.
- *fine-tuning the scheme:* the user can adjust the personalisation controls (7) to holistically modify, for example, the hues used and the saturation of the colour scheme. Any changes would be immediately reflected in the displayed interface colour scheme.

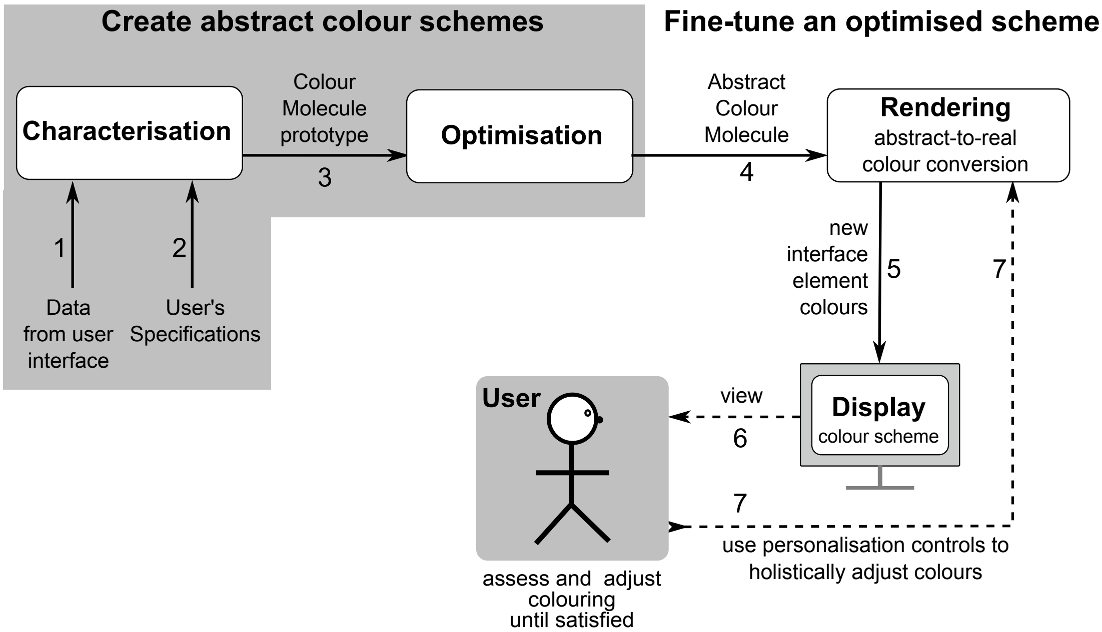


Figure 3.6: The proposed colour scheme design cycle, from the extraction of data from the user interface and specification of constraints by the developer, through to the creation of an optimised colour scheme, the colours of which are used to recolour the interface elements. The developer can holistically alter the abstract-to-real colour mapping to change the overall scheme colours to get the most pleasing result. Also illustrated is the flow of data within the Colour Harmoniser architecture. The dotted lines (6 & 7) are external: 6 is the user looking at and evaluating the displayed colour scheme, and 7 is the user adjusting the personalisation controls to holistically alter the scheme, which will cause the displayed colours to be immediately updated.

Because colour harmony is a holistic percept, it is desirable that the effect of the developer's adjustment of the scheme colouring should be visible immediately and in context. This can be achieved if the colour personalisation controls directly update the colours of the target interface. If the personalisation controls do not affect the relationships between the colour atoms, the integrity of the colour scheme will be preserved: it will continue to satisfy the design constraints, both while being adjusted and afterwards.

3.11 Summary

The design of a colour scheme using this approach would require the software or (where specified) the developer to:

1. *characterise the interface* by

   - extracting the areas of the interface elements from a definition of the interface to be coloured.

[12] "to tweak" – to fine-tune or finely adjust

- creating a list of interface element pairs that relate to readability, each pair being a text elements and its background.
- specifying the shape of the colour scheme wireframe to be used as the basis for the created scheme.
- defining the grouping of interface elements.

2. *specify the colouring constraints:* the developer would define any distinguishability constraints between the interface items to be coloured.
3. *create a set of preliminary abstract colour schemes:* by positioning some colour atoms randomly on the wireframe and some randomly in the colour space.
4. *optimise the abstract colour schemes* by
   - finding colour atom positions that satisfy the design constraints.
   - deriving real colours from the optimised abstract colour space coordinates.
   - updating the colours on the displayed interface from an optimised scheme.
5. *personalise the real colour scheme* by allowing the developer to change the actual colours used to colour the interface elements without affecting the integrity of the optimised colour scheme.

A system based on this model would

- *minimise the need for design-oriented specifications:* the developer will typically not have colour design expertise, nor the related vocabulary. By extracting as much data as possible from the interface itself, only the essential and non-specialist requirements need to be provided by the developer before initial scheme creation can begin.
- *create colour schemes with broad appeal:* colour preference is subjective, but there are broad principles that can be used to design colour combinations that are likely to have broad appeal.
- *ensure readable text within colour schemes:* there are criteria based on the lightness difference between text and its background that can be used to guarantee readability.
- *support the iterative nature of creative design:* artists and professional designers iterate when developing a creative work, modifying the work until it looks right. A system for colour scheme design should support the same iterative process, rather than excluding the developer from the design process. The developer may not know how to create a colour scheme, but will almost certainly have opinions on any created schemes.
- *allow for the variability in human colour preference and domain-specific constraints on colour use:* the completely automatic creation of a final colour scheme

is unlikely to be successful due to the need to match the colours to the personal preferences of the developer, and also to the context and culture of the intended audience. Such effects can be allowed for as part of the post-creation colour scheme adjustment process.

- *produce schemes usable in normal working conditions:* the combination of consumer standardisation on sRGB, and the adaptation inherent in human colour vision, as well as the empirical evidence from the widespread satisfaction with colour on the World Wide Web (as indicated by the acceptable colours of images from photo and video sharing sites[13]), indicates that colour schemes created with colours encoded to the sRGB standard would be seen as intended.
- *possibly produce colour schemes that are helpful to CVD users:* the lightness contrast used to ensure readable text would aid legibility for CVD users.

An implementation of a software tool based on the design outlined above is described in the next chapter.

[13] e.g. `http://www.flickr.com`, `http://picasa.google.com`, and `www.youtube.com` - all three URLs accessed May 21, 2010.

Chapter 4

The realisation of a colour harmoniser

The last chapter introduced the functional blocks that could be used to make up a system for automatic interface colouring. This chapter will describe one possible implementation of that architecture. The prototype has four major sections, mirroring those described earlier – characterising an interface, specifying the developer constraints, creating and optimising colour molecules, and finally, allowing the user to personalise the colours by exploring variations of the generated colour schemes. The prototype is called the "Colour Harmoniser".

Pictures of various colour schemes are included as illustrations in this chapter. Where one figure contains multiple schemes, the same underlying wireframe (e.g. split-complementary) will be used for them all. Then to illustrate the different types of colour schemes that are possible, successive examples will each use different wireframes (e.g complementary for the first, split-complementary for the next, complementary including black-white etc.). This will allow the images in each figure to focus on the intended point (rather than possibly distracting factors related to the usage or colours from the different wireframes), while still, by the end of the chapter, displaying a wide variety of different colour schemes.

4.1 Characterisation of an interface

To create a harmonious colour scheme for a specific interface, the interface must first be *characterised*. During this phase, interface characteristics such as the areas and types of elements, and whether they are shapes or text, are captured so that this information can be used later, to tailor the interface colour scheme to its characteristics.

This information must be extracted from whatever specification of interface is available. There are may different forms, including a design-time specification from a GUI software development environment, a context-specific web design tool such as Adobe *Dreamweaver*$^{\text{TM}}$, a GUI interface framework using Java, Javascript or Python, or a specification in which the interface element objects and their placement are coded textually (e.g. HTML).

If the interface is based on a document object model (DOM), the colourable interface elements that it contains can be found by programmatically querying the top-level interface object. The Colour Harmoniser prototype was developed in Borland's *Delphi*™ which provides this ability, as would Java, C# or any modern interface development environment, including one based on HTML and CSS (cascading style sheets).

4.1.1 Determining the area of the colourable interface elements

In a graphical user interface, it is common for the visible objects to be layered, with smaller objects being positioned on top of other larger elements: e.g. buttons on top of a panel, which is itself on top of a background frame. The buttons would be fully visible, but only part of the panel and part of the background remain unobscured. Any colour harmonisation algorithm using area as a parameter needs to know the visibly exposed area of each element, not its bounding box. Moreover, for the Colour Harmoniser to be practically useful, the visible area calculations of the colourable interface elements must be automatic.

The area of an interface element can be less than the area calculated from its $width \times height$ bounding box. Consider a background panel (e) behind non-overlapping foreground elements $f_1 \ldots f_n$. The visible area (A_v) of e is:

$$A_v(e) = A_t(e) - \sum_{i=1}^{n} A_v(f_i) \tag{4.1}$$

where:

$A_v(\cdot)$	the visible area of a single interface element
$A_t(\cdot)$	the total area – the area of the element's $width \times height$ bounding box
e	a background interface element
f_i	foreground interface elements: those in front of and obscuring e

To calculate the visible area of all the interface elements, their size, position, and Z-order (depth ordering) are required. Once the visual hierarchy of the elements has been established and the area of each element is known, the *visual area* or *visible area* of all the visible elements can be derived. If any of the elements partially overlap, the situation is more complex. However, this is much less common than nested elements and is not considered further.

Interface elements properties – programmatic considerations

The interface element properties available from a GUI design environment are not intended to provide ready access to colour and area properties. This causes complications in two ways:

finding the area: Determining the visible area of interface elements using equation 4.1 requires the depth ordering of interface elements to be known. This is not necessarily straightforward, as:

- the Z-order of objects returned by a programmatic query reflect the object hierarchy, not the visual hierarchy.
- interface objects used in GUI design environments often have Boolean properties such as *transparent*, *visible* or *enabled*, which allow the object to be listed in the object hierarchy without being visible (i.e. an object A can be "closer" to the viewer than object B without obscuring B).
- some elements may be partially transparent.

controlling an item's colour: access to the colour property of interface elements is needed in order to update the displayed colour to reflect a new colour scheme. However, while conceptually simple, in practice, accessing this property is also not straightforward because:

- some basic controls have more than one colour property (e.g. *textColor* and *backgroundColor*).
- programmatically searching for the colour property of visual components can be complicated by inconsistent naming conventions. In Delphi 7 for example, some controls have a *color* property, others have *BackgroundColor*, and in others the colour property is nested (e.g. *font.color*).
- some composite controls have a complex internal structure with many colour properties. For example, a 3D plotting control can have properties[1] like *3D-bar-top-color*, *bar-side-color*, *bar-front-color*, *background side-wall-colour*, *back-wall-colour*, and a colour for each plot series. These graphically complex controls were omitted from the trial implementation.

This list is not intended to be exhaustive, but illustrates some of the practical difficulties encountered in the conceptually simple task of automatically determining the properties of colourable interface elements.

To regularise access to the many differing colour controls and properties, the adaptor design pattern Gamma et al. (1994) was used. This provides a consistent programming interface to both the colour and area properties, and enables controls with n multiple colourable subitems to appear as n separate controls, each with its own *color* and *area* properties.

For the commonly used controls, the Colour Harmoniser prototype handles all of above cases except that of semi-transparency. As semi-transparency is not widely used in GUI and web design, its omission from the Colour Harmoniser prototype does not unduly restrict the generality of the colour schemes that can be designed.

Textual interface elements

There is one remaining complication – textual interface elements – whose visible area is always less than their bounding box and is a function of the text, the font and the font size and style. To determine the area of a string, it could be rendered and

[1]These names are illustrative, but are related to the property names in actual Delphi controls.

the total number of pixels occupied by the characters counted. However, there is no guarantee that the design-time text and the text that will eventually be displayed when the application is run or the web page is displayed are the same. The smaller elements (e.g. menus, captions, headings, etc) are less likely to change than large areas of text, such as those found in web sites with dynamic content and in desktop applications.

If the exact area of textual elements is needed, evaluation must be deferred until the actual text to be displayed is known. However, it may not be desirable to dynamically adjust the area of the text as this could alter the colour scheme, for reasons that will be explained later in this chapter. Instead of measuring the textual pixels, an estimate based on the textual element, (e.g. bounding box, its font/size/style (bold, italics)) could be used. Simpler still is an estimate that assumes that the text covers a predefined fraction of the area, as is done by printer manufacturers when estimating the life of their printer cartridges. Text is assumed to cover a fixed percentage (e.g. 5%) of the area of each page. The Colour Harmoniser prototype uses the same technique, but uses 10% instead. This higher value is based on measurements of two test interfaces – a prototypical web page and a presentation slide. Unlike a printed page, interfaces do not have a large blank border around the text. The mean difference between the measured number of pixels and the value obtained using 10% of the bounding box area of the element (i.e. the button or body-text background) was ∼3%. A more complex function may be justified for a production system. The form of this function would need to be determined experimentally, but for the user trials with static text in a fixed font, the use of a fixed scaling factor appeared satisfactory.

4.2 Including developer-specified colouring constraints

Once a list of all the colourable items on the interface is available, the developer may wish to specify how these individual elements should be treated – whether they are to be treated as individually colourable elements or grouped to reflect the semantics of the interface.

4.2.1 Colour and user interface semantics

The colour choice for many of the interface items will be unconstrained. However, the developer may wish to specify sets of items that should be the same colour, and pairs of items that should *not* have the same colour. These form sets of items that should be visually distinct[2]:

- to ensure visibility: if one item sits on another, the smaller one would be invisible unless it is a (significantly) different colour from the larger surrounding element.
- to ensure that an end-user doesn't assign unwarranted behavioural expectations (e.g. related functionality) to controls with identical or indistinguishable colours.

[2] *Visually distinct* does not just mean *not the same colour*. Two colours may be technically different but visually indistinguishable. To be visually distinct, two objects must be sufficiently differently coloured for the distinction to be immediately apparent.

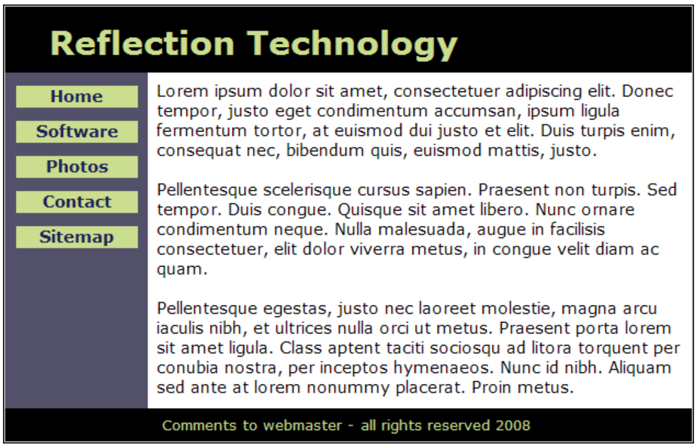


Figure 4.1: A prototypical web page.

- to satisfy the developer's aesthetic sense: the developer may simply want two objects to be coloured differently.

For example, on a typical web page (fig. 4.1), the navigation buttons perform a similar task and are often coloured identically. To reflect this, they could be grouped into a "Buttons" meta-element (group). To use the same colour for the header and footer backgrounds (as is also common), they could be put into "HeaderAndFooterBgnd" group. The developer may decide, for aesthetic reasons, to require that the "Buttons" and "HeaderAndFooterBgnd" should be coloured differently. Any of the elements (e.g. body-text-background, header and footer text, the navigation bar) could be required to be distinct from any other individual or grouped element.

The auto-classification of interface elements

An interface of any realistic complexity can contain a large number of colourable items (e.g. 17 in fig. 4.1; for details, see Table 5.3 on p170). The developer may want to partition the interface elements into semantically related groups, so the colours of the elements reflect the interface semantics, as discussed in section 3.1.

It is simpler for the developer if some initial groupings happen automatically, as this reduces the number of items that the developer has to classify. In the early development of the Colour Harmoniser prototype, an automatic classifier was built that created groups with a structure based on the object hierarchy of the interface. For example, a panel containing buttons would result in a group for the panel, with members of the group being the buttons[3]. This was not as useful as was hoped, as real interfaces often use a large number of visual container objects, many of which are completely obscured and exist solely to control the alignment of the contained items. Basing the group creation on the object hierarchy can result in a large number of automatically created groups, many of which are of no interest from the colour design perspective. More awkwardly, interface elements that are semantically equivalent (buttons or text

[3] or more correctly, the members of the group were colour atoms, one per button.

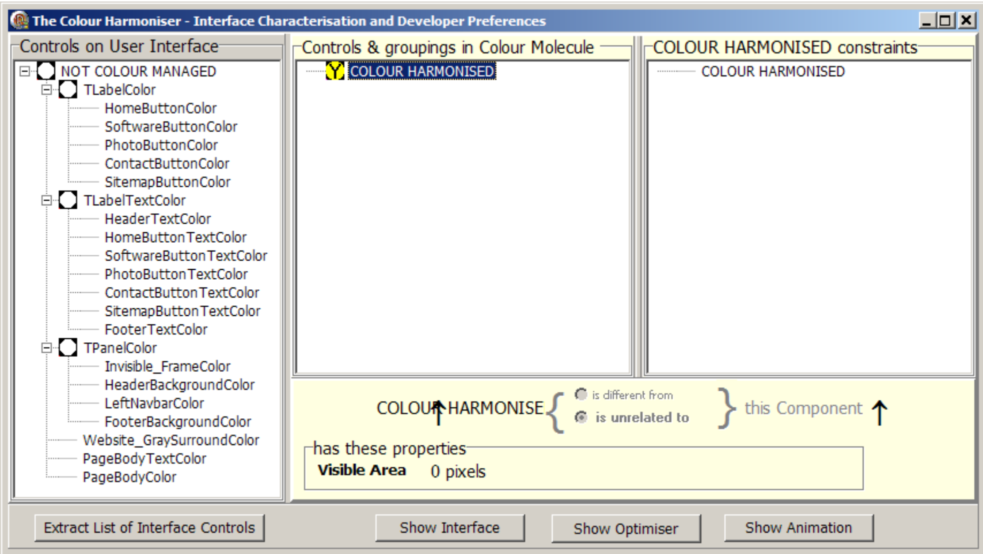


Figure 4.2: The left hand panel shows the interface element names extracted automatically from the web interface (fig. 4.1). Multiple elements of the same type are placed into an automatically created group (named using the underlying class names, which unfortunately are not usually self-explanatory). In the interface, the textual elements are in several different containers, but the automatic classifier groups them together, e.g. "HeaderTextColor" and "HomeButtonTextColor", which are items of text have been placed together into the automatically created group "TLabelTextColor", and items of class "TLabelColour" (the rectangular backgrounds to the text) have also been grouped. Unfortunately, the structure of the automatically created groups does not match the visual structure of the web page, but this can be remedied by renaming the groups and regrouping elements (fig. 4.3).

for example), will be placed in different groups if they happen to be in different panels on the interface. Without manual regrouping, there was no easy way to indicate that all similar items (e.g. text) should be coloured identically.

The automatic classifier was altered to create groups based on the class of the interface objects, not their parent in the object hierarchy and so, for example, all textual element are placed in one group, all buttons in another. This was found to be simpler to use. There were fewer automatically created groups, and the groupings seemed more natural and required fewer alterations. The extraction of elements from the web page (fig. 4.1) is shown in the left-hand panel of figure 4.2[4].

4.2.2 Meta-objects and identically coloured items

The holistic approach to colour harmony does not allow the designer to specify the colour of an individual object. It does, however, allow the designer to specify that objects (or sets of objects) should have the same or different colours. Such a constraint

[4] not all the backgrounds to the textual elements are visible (e.g. as HeaderTextColor's background is transparent, it is omitted from the list of elements).

could be handled during the optimisation by forcing the colours of members of an identically-coloured set together in the colour space, and by forcing the colours of the members of differently-coloured sets apart.

In the case of identically coloured objects, this would be highly inefficient as it would require significant unnecessary calculation that could be circumvented by grouping objects, (hierarchically if necessary), into meta-objects. The area of a meta-object is the aggregate of the areas of its member objects, and any colour assigned to the meta-object is propagated to its member objects. A group member may be an individual element or another meta-object[5]. Grouping the objects before the optimisation phase significantly reduces the number of objects to be considered and thereby allows the optimisation to run more quickly, so the grouping of interface elements was included in the Colour Harmoniser prototype. There is one caveat to this simplification. When seen from a normal viewing distance, the interface elements being grouped must be sufficiently large to appear as discrete elements and not to form a region of dithered colour. For normal-sized interface elements, this is not a problem.

After the definition of the same-coloured groups, the area of any single interface element or same-coloured group may be found using the augmented *VisibleArea* function A:

$$A(e) = \begin{cases} A_v(e) & \text{for an individual interface element } e \\ \sum_{i=1}^{n} A_v(e_i) & \text{for an interface element group with members } e_1 \ldots e_n \end{cases} \tag{4.2}$$

where:

$A(\cdot)$	a function that returns the visible area, allowing for layering and grouping
$A_v(\cdot)$	returns the visible area of a single element (equation 4.1).
e, e_i	an interface element

The Colour Harmoniser prototype allows the developer to select the elements to be part of the colour molecule by dragging them from the left-hand panel into the centre panel (see fig. 4.3. This selection allows some elements to be omitted from colouring if desired. The contents of the right-most panel will mirror the contents of the centre panel for reasons that will be explained presently. The elements in the centre panel can be moved to form "same-coloured" groups. In the demonstration interface, there are four sets of elements that have the same colour: the header/footer text, the header/footer background, the text on the buttons, and the rectangular buttons themselves. After the creation of several user-defined groups and the renaming of elements to clarify the intent, the element groupings reflect this structure (fig. 4.3). The yellow "Y" icon at the top of the centre panel indicates the colour scheme wireframe to be used, in this case, the Y-shaped split-complementary scheme. The right-click popup-menu allows the scheme to be changed and provides access to other options (fig. 4.4).

[5] an architecture that corresponds to the composite design pattern Gamma et al. (1994).

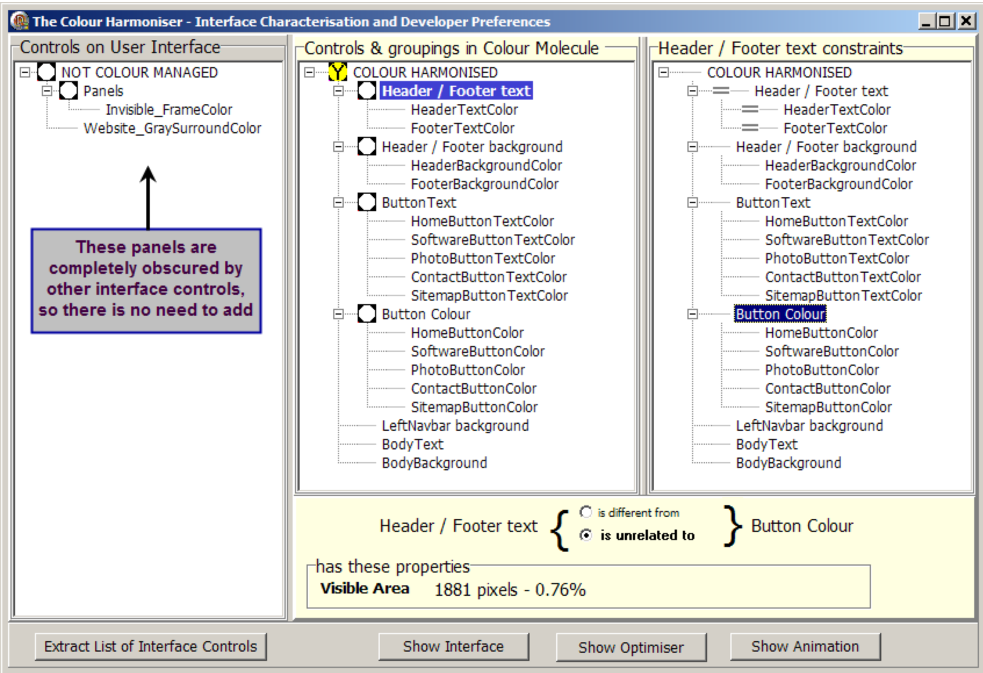


Figure 4.3: The user indicates those interface elements whose colours are part of the colour molecule by dragging them from the left-most panel to the centre, where they can be grouped into a hierarchical structure and given more meaningful names. The Y-shaped icon with yellow background indicates the colour scheme wireframe ("Y" denotes split-complementary), and the octagonal black-edged icons indicate "same coloured" groups. The right-most panel mirrors the centre panel. Its use is explained in figure 4.5.

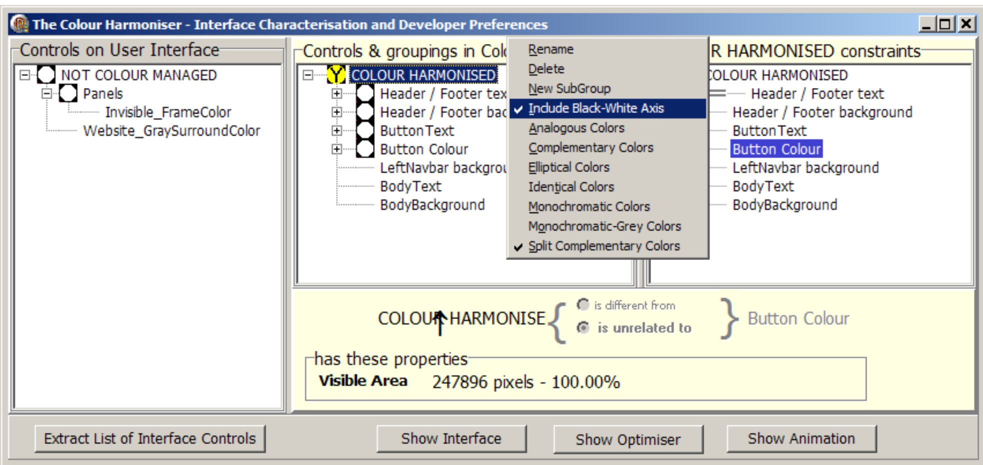


Figure 4.4: A popup menu can be used to create new groups, to rename elements, and to specify the wireframe to be used. If desired, the black-white axis can be included as part of the wireframe.

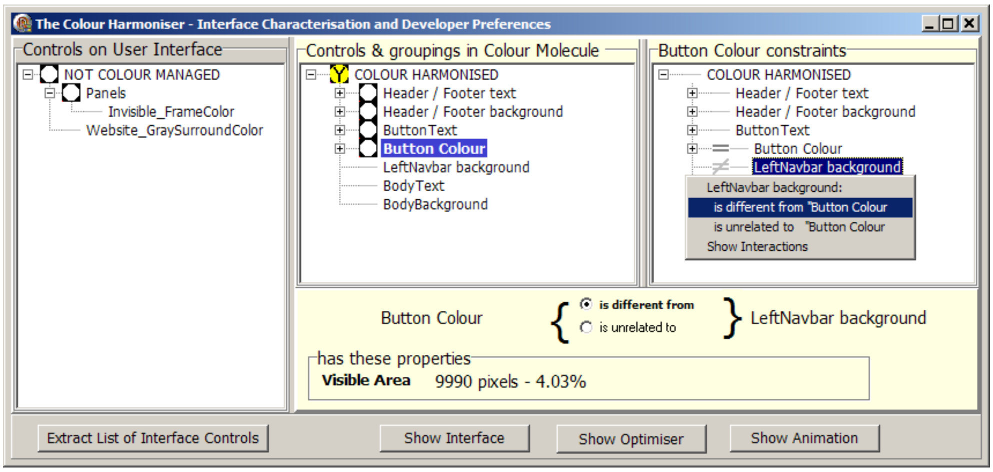


Figure 4.5: The groups can be collapsed to hide unnecessary detail. Distinguishability constraints can be defined by first selecting an item in the centre panel (e.g. *Button Colour*), then right-clicking on an item on the right-hand panel (e.g. *LeftNavBar Background*) and selecting the appropriate constraint from the popup-menu. As shown above, the "is-distinct-from" is selected, meaning the *Button Colour* should be coloured differently from the *LeftNavBar Background*. Any constraint between the two selected items is also shown in the right-hand panel: the "$\neq$" icon indicates "is different from", "=" is used to indicate the elements have the same colour (are part of a group), and no icon means there is no colouring constraint between the selected elements. Instead of using the popup menu, the radio buttons below the centre and right panels can also be used.

4.2.3 Distinguishability constraints

The developer can specify that any pair of interface elements must appear distinguishable (fig. 4.5 and 4.6). Unfortunately, simply checking that the specified pairs of elements are distinguishable is insufficient, as it does not address a subtle complication that is caused by the introduction of the same-coloured groups. To allow for the effects of grouped elements, the developer-specified distinguishability constraints and the definitions of same-coloured groups are used to define a Boolean *Distinguishability-Constraint* predicate ($d(a, b)$) is *true* if the developer has specified that groups a and b must be distinguishable (where either group may consist of a single element). The developer may also specify a distinguishability constraint against a member of a same-coloured group, in which case, that constraint is automatically propagated to all the other members of the group.

Developers can specify the desired colouring relationships using the distinguishability constraints and the composition of the same-coloured groups, which can then be used to ensure that the colour relationships in automatically created schemes reflect both the interface semantics and the developer's aesthetic preferences.

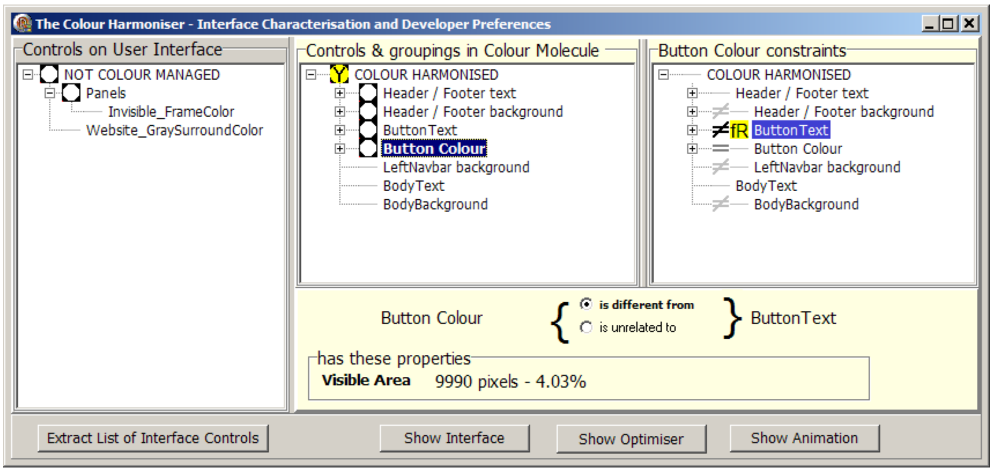


Figure 4.6: The constraints shown will force the colour of a button to be different from the header/footer background, the left navigation panel, and the main body background. The "fR" (in the right-hand panel) indicates an automatically defined "forced" readability constraint between the button and the button text.

4.3 Ensuring legibility of the coloured interface

A colour scheme may be aesthetically pleasing, but unsuitable for use in an interface, if any text is difficult or impossible to read. Text/backgound pairs can be detected automatically, and an obvious way to ensure that the text remained legible would be to add these pairs to the list of user-specified distinguishable items. Unfortunately, this appealing simple approach will not always work, because the distinguishability constraint is based on colour space separation in any direction in the colour space (measured by ΔE), whereas to ensure readability requires a more severe constraint, a guaranteed difference in lightness ΔL (see section 2.9.8).

Ensuring the readability of text is analogous to imposing a user-defined distinguishability constraint. The constraint checking is complicated by the need to allow an item of text and/or its background to be a member of a same-coloured group. A Boolean Readability Constraint predicate (r) indicates those pairs of items that must be given appropriate differences in lightness by the optimiser.

$r(a, b)$ is true if any of the following are true:

- t is an item of colourable text, and b is its background, OR
- t is an item of colourable text and b is a same-coloured group where $\texttt{background}(t) \in b$, OR
- t is a same-coloured group containing an item of colourable text T and b is a coloured interface element where $b = \texttt{background}(T)$, OR
- t and b are both same-coloured groups and if an item of text t_i is a member of group t, and its background b_j is a member of group b

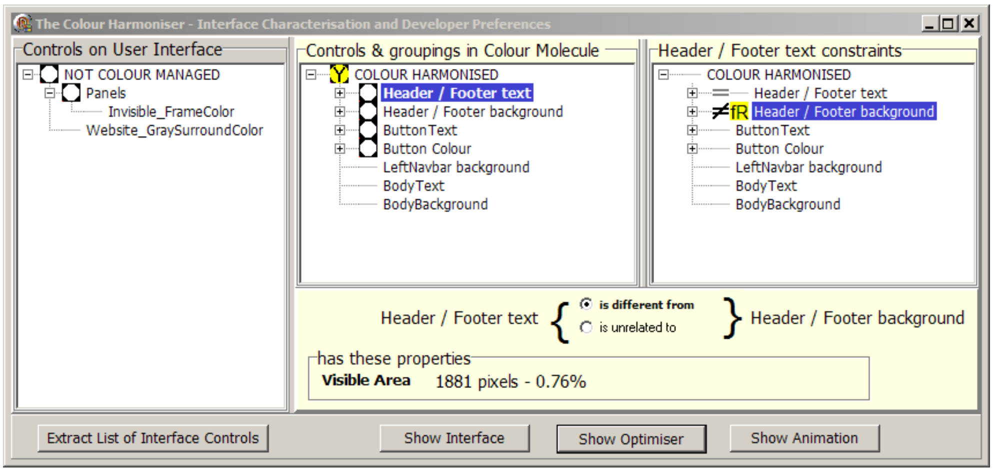


Figure 4.7: Once the interface has been characterised, the interface colouring is determined from the properties of the "top level items" – the children of the "Colour Harmonised" parent. To ensure constraints are handled correctly, distinctions between any children (e.g. footer-text and footer-background) in different groups require the constraint be also applied between the user-defined groups (Header/Footer Text and Header/Footer Background).

(a readability constraint between any member of t and any member of b must cause the groups themselves to have sufficient δL difference).

Figure 4.7 illustrates the forced readability constraint between the group containing the header/footer text and the group containing the header/footer background elements. As the text must be different from the background, and all elements in each group are the same colour, the groups themselves must be different.

Once the Harmoniser has extracted the interface elements and the user has indicated their preferred colouring; by grouping elements, by defining distinguishability constraints, and by selecting their preferred colour scheme, the following information is available:

- a list of colourable items on the interface and a way of finding their visible area, and setting the colour of any item. The items may be either individually colourable interface elements or same-coloured groups.
- a Distinguishability Constraint predicate (d) – a method of determining, for any pair of items[6], whether or not the developer has specified that they should be visibly distinguishable.
- a Readability Constraint predicate (r) – a method of determining, for any pair of items, whether or not they should be coloured to ensure readability.
- the wireframe that the developer has chosen as the basis for the colour scheme.

[6]Either of the arguments to the Distinguishability or Readability Constraint predicates can be a single item of a single-coloured group.

- whether or not the colours for the interface objects may include achromatic (black-grey-white) colours, in addition to those along the wireframe path.

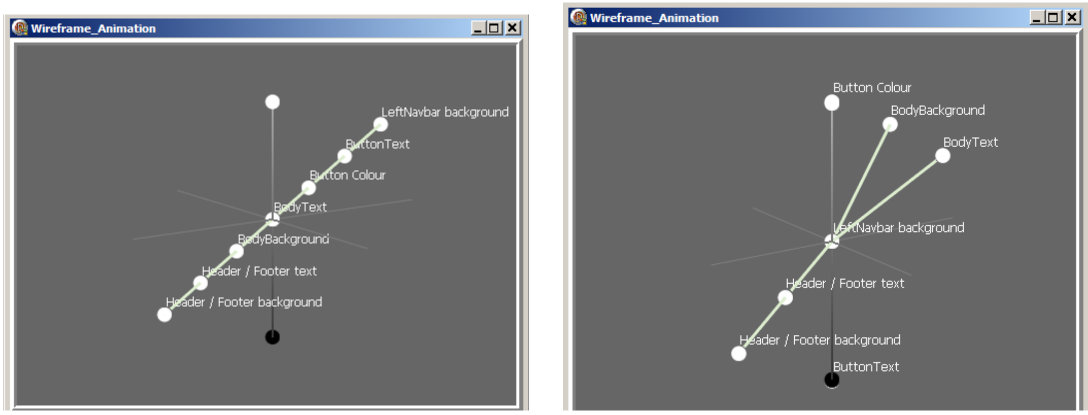


Figure 4.8: Two initial arrangements of colour atoms on different wireframes. On the left, a complementary wireframe with the items equally spaced. On the right, a split-complementary wireframe extended to include the black-white axis as part of the wireframe.

4.4 The creation of an abstract colour molecule

Once the interface has been characterised and the developer has specified a set of colouring constraints and a wireframe archetype, it is possible to allocate a position on the wireframe for each interface element. If the maximum length of a wireframe segment from the origin is ± 1, and if any orientation of the wireframe is possible, then only points within a unit sphere can occur within a colour scheme. Limiting randomly selected colours to the points within this sphere limits distances is useful, as it limits the maximum distance in the abstract colour space.

4.4.1 Creating prototypical colour molecules in abstract colour space

The interface characterisation phase allowed the developer to indicate the number and composition of the colour atoms (elements or groups) in the interface. The positions of the colour atoms on the wireframe defines the lightness and saturation of the related elements, and therefore defines the colour relationships between them. There is a large number of possible positions. As an initial allocation, the atoms could be positioned randomly on the wireframe, or fixed points could be chosen and the colour atoms allocated to these points randomly.

The method used in the Colour Harmoniser prototype when initially creating colour molecules is to randomly allocate atoms to a fixed points chosen to be as far from each other as possible. This maximizes the average colour difference between adjacent atoms, and, for some orderings, will result in distinguishable (or readable) interface elements. The method used to choose the fixed points depends on the wireframe. For single line wireframes, the atoms are evenly spaced, as shown in figure 4.8.

For wireframes composed of multiple straight-line sections (e.g. split-complementary), the atoms are allocated to each segment in turn. For curved wireframes, the points are allocated at equal rotational angles from the centre. The actual distances will vary, but precision is not necessary, as the atom positions will be altered later to find an arrangement that satisfies all the constraints. Figure 4.9 illustrates the full set of wireframes supported by the Colour Harmoniser prototype, showing an initial set of positions and the placement of the wireframe in the abstract 3D colour space.

It is very unlikely that any of initial orderings will exactly satisfy the colour strength balance and readability/distinguishability constraints, but some orderings will approach this ideal more closely than others. For example, some orderings will allocate a textual item and its background to widely separated points on the wireframe. If the wireframe is tilted (e.g. to 45° to the vertical), more widely-separated wireframe positions will have greater lightness difference and are more likely to satisfy the readability criterion than orderings in which the atoms are closer (and the ΔL is correspondingly lower). The evaluation of distinguishability is similar to that of readability, but is evaluated from the less restrictive ΔE (rather than ΔL) separation, so more separated items will have better distinguishability, and for colour balance, some orderings will be more balanced than others. To satisfy all the criteria at the same time, it is very likely that the atoms will need to be moved from their initial positions.

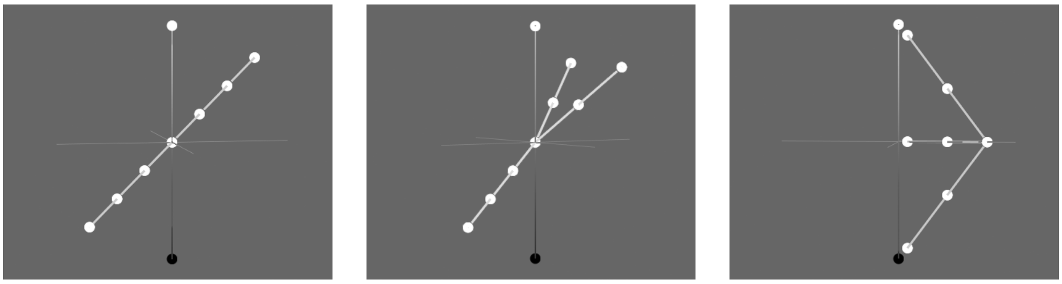
4.4.2 Augmenting the wireframe with achromatic colours

The wireframe upon which a colour scheme is based strongly affects the range of possible colours. Monochromatic colour schemes use one hue; complementary schemes use two; split-complementary schemes use three; less commonly used wireframes, such as those based on curves (e.g. circular/elliptical) use many.

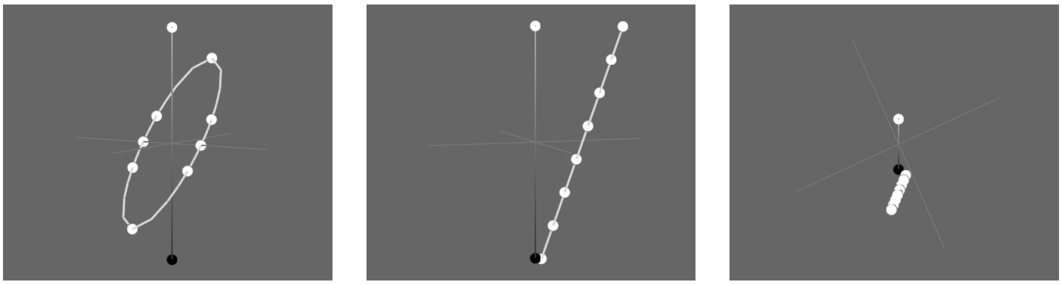
In the initial development of the prototype, the schemes produced were colourful and balanced, but often too colourful for a graphical user interface. It wasn't that colour was being used badly: there was just too much of it – everything was coloured. This overuse of colour is not particular to interface colour schemes, it also applies to art: *"color can overwhelm... one must understand that when it comes to color* less *is often* more *– a lesson taught us by the masters but ignored by many artists"* – Singer (1976), cited in Edwards (1993). The primary role of an interface is to convey information, and the use of too much colour can be distracting. It is therefore desirable to provide a means of moderating the use of coloured elements. Comparing schemes on the web and desktop GUI interfaces with those produced by the prototype, the reason for the difference easy to see: black, white and shades of grey were missing from the schemes produced by the prototype. These achromatic shades are widely used in real interfaces, but cannot (with the exception of mid-grey) occur in a colour scheme based purely on any of the common wireframe shapes.

To address this, an option was added to allow blacks, whites and different greys to appear in the final colour scheme. This is implemented by extending the wireframe to include an additional path, along the lightness axis[7]. This is a simple but significant augmentation of the wireframe archetype. It allows the user to choose (for example) a complementary colour scheme enhanced by the addition of black, white and grey elements, instead of a scheme consisting totally of complementary colours. This modifi-

[7]the vertical (y axis) in the abstract colour space is mapped to lightness in the abstract-to-real colour space mapping.



Complementary | Split-Complementary | Monochromatic



Elliptical | Analogous – side view | Analogous – top view

Figure 4.9: The colour scheme wireframes supported by the Colour Harmoniser prototype. The top row shows the wireframes for most frequently used colour schemes: complementary, split-complementary, and monochromatic (where only one hue is used). The monochromatic wireframe includes a pragmatic optimisation: it stops short of the black-white axis, so that the lightest and darkest elements contain a touch of colour, resulting in a more sophisticated looking scheme. On the bottom-left is the elliptical path suggested by Munsell. The bottom-centre and bottom-right figures show the analogous wireframe used in the prototype. The analogous scheme is often shown as an arc on a full saturation colour wheel, but this is too restrictive for use in the prototype: changes in lightness are obtained by tilting the wireframe from the horizontal; changes in saturation by using a different radius for the top and the bottom of the wire, as shown in the bottom-centre figure; and changes in hue by tilting the wire from the vertical, which changes the rotational angle for the different atoms, as shown in the bottom-right figure.

cation can be applied to any of the balanced (complementary, split-complementary and elliptical) wireframes. With the addition of the black-white axis, the resulting schemes more closely approximate many of the more visually appealing interface colour schemes used on websites and in desktop applications.

The use of achromatic colours does not appear to disturb the colour harmony, and, by simplifying the scheme, accentuates those elements that are coloured.

4.5 The algorithmic evaluation of colour schemes

The purpose of the Colour Harmoniser is to produce colour schemes that are at least acceptable and hopefully good. Given the degree of subjectivity in human colour scheme preference, a certain amount of variability is to be expected, but nevertheless, for the Colour Harmoniser to succeed, it must be able to assess an interface colour scheme in a way that correlates strongly with human assessment of the same scheme.

A colour molecule represents an abstract colour scheme, but there is a very large number of schemes and no deterministic method of creating good ones. Therefore, it is necessary to use an iterative process, a search, which will require some method of comparing colour schemes – a fitness or objective function – that assesses the factors related to the use of colour within an interface.

The fitness function used in the Colour Harmoniser prototype calculates a single numeric score that reflects how well a colour scheme satisfies the usability constraints defined by the developer and the colour aesthetics, using models and heuristics from the artistic community, augmented where necessary for use with user interfaces. If the score for a colour scheme correlates well with the human perception of the same colour scheme, then the fitness function may be used to evaluate the appeal of any scheme represented as a colour molecule, and may be used in an optimisation process to improve the suitability of a colour scheme for interface use.

The fitness function only assesses those factors relating to colour harmony, readability, and the scheme's conformance with the developer-specified constraints on colour difference. More subtle factors such as: "equilibrium, symmetry, sequence, cohesion, unity, proportion, simplicity, density, regularity, economy, homogeneity, and rhythm"[8], that pertain to the aesthetics of the complete interface, will already have been considered by the designer of the interface, and are not considered here.

In the discussion so far, the colour atoms have been assumed to be constrained to lie on the wireframe. This is the ideal situation, but it may not be possible to simultaneously satisfy the aesthetic criteria and the readability and distinguishability constraints while constraining items to the wireframe. Therefore, to allow some flexibility in the constraint resolution, colour atoms may move off the wireframe, but it is desirable for them to remain as close to the wireframe as possible, so a "wireframe alignment" factor has been included as part of the fitness function. This factor returns an optimal score when all the colour atoms lie on the wireframe, and a decreasing score as one or more atoms move away from the wireframe.

As the schemes are no longer restricted to the wireframe, it is possible for the fitness function to return a score for a set of colour atoms with arbitrary positions in the colour space. This enables a fitness score to be derived for molecules with randomly positioned atoms, or molecules with any blend of random and wireframe-aligned atoms. This results in more colour scheme variations and a greatly expanded the search of the solution space; this topic is discussed further in section 4.6.

The Colour Harmoniser therefore incorporates a fitness function that evaluates four

[8] This list is derived from Ngo et al. (2002), which details of an algorithmic approach to the assessment of these factors, and the derivation of an *order–complexity* scale.

separate contributions and combines them to give an overall score for an abstract user interface colour scheme. Of the four, the first two relate to aesthetics, the latter pair to usability. In outline, the four components are:

colour strength balance: an assessment of whether, when area is taken into account, the colour scheme is well-balanced around mid-grey. Monochromatic schemes, by definition, do not balance around grey, so the inverse-area rule (sec. 2.9.6) is used instead.

wireframe alignment: an assessment of how the positions of the colour atoms conform to the wireframe path. The inclusion of this term allows an evaluation of a molecule with an arbitrary set of colour atom positions.

distinguishability: an assessment of whether the interface elements the developer has specified as distinguishable are sufficiently separated in colour space.

readability: an assessment of the readability of the text in an interface.

4.5.1 Normalising the fitness term scores

Each of the four contributions to the fitness value of a colour molecule will be normalised to fit within the range from zero to one; in each case, a value close to one indicates a good score for that term, and a value close to zero indicates a low score. The measurements upon which these contributions to the score are based do not conveniently fit into these ranges, and in two cases increase in the direction of low colour harmony, so a normalisation function is applied:

$$N(x) = 10^{-x} \tag{4.3}$$

where:

x	–	the value to be normalised, where $x \geq 0$
$N(x)$	–	returns a value in the range 0 to 1.

The function of the form n^{-x} is arbitrary, but has two useful properties. Firstly for a given range of x, the rate of falloff is controllable by modifying n. Secondly, for positive inputs, it returns a value in the range 0 to 1, giving a well defined (and equal) range to each term of the fitness function. The values from the various fitness subfunctions have a minimum value of zero, but with ill-defined maxima. A power function like n^{-x} transforms these values, so an input values of zero return one, and increasingly positive values return values that fall, but remain greater than zero.

The normalising function does introduce non-linearity into the fitness function but, even without the inverse power function, there is no reason to believe that a scheme with a fitness score of 0.5 would appear in any meaningful way to be "twice as good" as one scoring 0.25. The fitness score is to enable schemes can be compared during an optimisation phase, not to measure the "quality" of a colour scheme in any absolute sense. It is only necessary that the function be monotonic and that higher fitness scores indicate better colour schemes.

4.5.2 Colour strength balance contribution

This term measures how well the current set of colour atom coordinates balance lightness and saturation against area in an attempt to have the average of both at mid-grey (CIELAB L = 50, a = 0, b = 0), or for wireframes not symmetric about the lightness axis, how well the inverse area criterion is met (sec. 2.9.6).

The function used is similar to, but not the same as, that resulting from Munsell's theory of colour harmony. The function evaluates balance about mid-grey, but evaluates area–lightness balance and saturation–area balance separately. For an image with two coloured components, Munsell's colour harmony theory balances two products:

$$A_1 \times V_1 \times C_1 = A_2 \times V_2 \times C_2 \tag{4.4}$$

where:

A_i	the area of an interface element
V_i	the value (lightness) of the interface element colour
C_i	the chroma (saturation) of the interface element colour

This allows an element's lightness and saturation to be traded off against one another. A generalised version of this function to handle multiple interface elements was initially implemented, but this was later modified to balance lightness and saturation separately, to allow the balance around mid-grey to be evaluated in the absence of a particular wireframe, and to allow for the inclusion of elements on black–white axis. It is implicit in Munsell's theory that the two colours being balanced are on opposite sides of the vertical axis, but this does not generalise easily to the different wireframe shapes. More importantly, if the colour strength balance can only be evaluated correctly if the wireframe is present, this makes the colour strength balance dependent on the wireframe alignment. It is desirable for the terms to be independent in order to make it possible to conduct an independent evaluation of the optimal weighting of each term in the fitness function from experimental data. It was therefore decided to balance lightness and saturation separately. This allows the evaluation of colour balance without any reference to a wireframe, and therefore removes the dependency.

Colour balance around mid-grey is calculated where possible. Mid-grey corresponds to the origin (0,0,0) in the abstract colour space[9]. However, for balance around the origin to be possible, a wireframe must pass through or enclose the origin. This is true for complementary, split-complementary, and elliptical wireframes, but not monochromatic or the related analogous scheme. Therefore, differing methods are used for balanced and unbalanced wireframes. For balanced schemes, a vector sum is used, balancing all

[9] the y axis corresponds to lightness, so (0,-1,0) is black, (0,1,0) is white, and the x, z plane corresponds to the chromaticity plane.

three colour space dimensions (in the range ±1) weighted by the area. For unbalanced schemes, the inverse area rule, which trades off area against saturation is used. The result is colour strength balance (C):

$$C(m) = \begin{cases} c_b & \text{– for wireframes encompassing/including the origin} \\ c_u & \text{– for wireframes offset from the origin, e.g. monochromatic} \end{cases} \tag{4.5}$$

$$c_b = \begin{cases} N\left(\sqrt{b_x^2 + b_z^2}\right) & \text{– if wireframe includes Black-White axis} \\ N\left(\sqrt{b_x^2 + b_y^2 + b_z^2}\right) & \text{– classical wireframes only, no B/W} \end{cases} \tag{4.6}$$

$$c_u = \frac{1}{n}\sum_{i=1}^{n} |d(a_i) - s(a_i)| \tag{4.7}$$

$$b_x = \frac{1}{n}\sum_{i=1}^{n} a_{i_x} A_v(a_i) \quad \text{– the balance of molecule } m \text{ in the } x \text{ dimension} \tag{4.8}$$

$$b_y = \frac{1}{n}\sum_{i=1}^{n} a_{i_y} A_v(a_i) \quad \text{– the balance for } y \tag{4.9}$$

$$b_z = \frac{1}{n}\sum_{i=1}^{n} a_{i_z} A_v(a_i) \quad \text{– the balance for } z \tag{4.10}$$

$$s(a_i) = \sqrt{a_i(x)^2 + a_i(z)^2} \tag{4.11}$$

$$d(a_i) = 1 - A_s(a_i) \times 0.9 \quad \text{– see notes below for details on the 0.9.} \tag{4.12}$$

where:

$C(\cdot)$	–	the colour balance score for molecule m
m	–	a colour molecule with n atoms: $a_1 \ldots a_n$
$c_b(\cdot)$	–	colour balance score for schemes *balanced* around mid-grey
$c_u(\cdot)$	–	colour balance score for *unbalanced* schemes, using inverse area rule.
$s(\cdot)$	–	a function that returns the saturation in abstract colour space
$A_v(a)$	–	returns visible area of atom a as a fraction of total visible area.
$d(\cdot)$	–	desired saturation according to inverse area-saturation guideline
$A_s(\cdot)$	–	a function giving the scaled area, 1 for largest element, 0 for smallest
$N(\cdot)$	–	the normalisation function – equation 4.3)

Notes:

$c_b(\cdot)$ – **the Black-White wireframe extension:** interface colour schemes frequently have large areas of achromatic colours. To allow this, the developer can choose to add the Black-White axis as a pseudo-wireframe element (sec. 4.4.2. If this option is used, the balancing on lightness is disabled, leaving only the saturation vs. area balance operational, for reasons discussed in section 4.5.3.

$d(\cdot)$ – **desired saturation:** The 0.9 in the definition of $d(a_i)$ limits the saturation of very large areas when using the inverse-area rule to 0.1 in the abstract colour space, rather than very close to zero. This gives a hint of colour to large areas and adds a degree of subtlety to the resultant schemes.

4.5.3 The black-white wireframe option and disabling lightness balance

Large areas of white are common on web sites and desktop application interfaces, without all the other elements being very dark, so it is clear that many widely used and acceptable interface colour schemes do not balance the lightness of the interface elements.

The inclusion of the black–white axis into the wireframe is intended specifically to allow the creation of schemes with large black or white areas. However, the inclusion of such elements makes the disabling of lightness balance desirable. Otherwise, balancing a large very light (or dark) area would require most or all the smaller elements to be very dark (light), or, to reduce its impact on the colour balance, the large area would be set close to mid-grey. The first of these options removes a large degree of flexibility in the allocation of colours (and is not consistent with the maxim that, in colour schemes, "black, white and grey go with anything"), and the second is the opposite of the intended effect (it would produce large areas of grey instead of large white or black areas).

Therefore, to allow the use of large white or black elements without significantly impacting the colour balance of the remaining elements, if the "allow black-white" option is in effect, the area vs. lightness balance is disabled, leaving only the area-vs-saturation balance in effect.

The importance of saturation balance and the lesser need for lightness balance is supported by the results of Morriss and Dunlap (1987) who found that users balanced chroma far more consistently against area than they did value, and that the relationship between value and area was more complex than it was between chroma and area.

4.5.4 The wireframe alignment contribution

The wireframe alignment calculation produces a score in the range 0 to 1 that is the mean average over all the atoms in the colour molecule of their individual wireframe alignment values. Those individual values are 1 for atoms that are exactly on the wireframe, and decrease exponentially as the atom moves further from the wireframe. This contribution to an abstract colour scheme's fitness value is thus a measure of the overall alignment of a set of colour atoms to the target wireframe. A score of 1 is achieved if all atoms lie on the wireframe.

$$W(m) = \frac{1}{n}\sum_{i=1}^{n} N(w_i) \quad (4.13)$$

$$w_i = |\, a_i - p(a_i)\, | \quad \text{– distance from atom } a_i \text{ to wireframe} \quad (4.14)$$

where:

$W(m)$	–	the wireframe alignment score for a colour molecule m
m	–	a colour molecule with n atoms: $a_1 \dots a_n$
$p(a_i)$	–	returns the point on the wireframe closest to atom a_i
$N(\cdot)$	–	the normalisation function – equation 4.3

High scoring molecules will therefore be those with atom positions on, or close to, the wireframe. As one or more atoms are positioned farther away, the score will fall.

4.5.5 The distinguishability contribution

The distinguishability calculation produces a score in the range 0 to 1 that is an average over all pairs of items that the designer has identified as needing to be distinguishable.

To evaluate how distinguishable one item is from another, it is necessary to know the real colours that will be associated with any abstract colour atom coordinate. For this to be possible, the transformation that will be used to map from the abstract colour space onto a real colour space must be known. Using this transform and the coordinates of any pair of items, it is possible to calculate their real colour difference.

The function used in the Colour Harmoniser prototype evaluates the distinguishability of a pair of items in two steps. It first calculates a similarity score in the range 0 to 1. The similarity between two items is inversely related to their colour separation, falling off exponentially as their separation increases. A distinguishability score between a pair of items (d'[10]) is inversely related to similarity:

$$\begin{aligned} s_{i,j} &= N(|a_i, a_j|) \\ d'_{i,j} &= 1 - s_{i,j} \end{aligned} \tag{4.15}$$

where:

$s_{i,j}$	–	similarity between items i & j, where 0 = no similarity
$d'_{i,j}$	–	distinguishability of items i & j: $0 \leq d' < 1$
$\|a_i, a_j\|$	–	Euclidean distance in abstract colour space between atoms a_i, a_j
$N(\cdot)$	–	the normalisation function – equation 4.3

The distinguishability value D' for a complete molecule m is calculated by averaging the individual distinguishability scores over all the item pairs that must be distinguishable:

$$D'(m) = \frac{1}{k} \sum_{i=1}^{n} \sum_{j=i}^{n} d'_{i,j} \quad \forall i, j \text{ where } d_{i,j} = \text{true} \tag{4.16}$$

[10] d the distinguishability constraint predicate, defined on page 119.

where:

k	–	is the number of pair of items that should be mutually distinguishable.
n	–	the number of colourable interface items.
$d'_{i,j}$	–	distinguishability of items i & j: equation 4.15
$d_{i,j}$	–	the distinguishability constraint predicate, defined on page 119

D' is an assessment of how well the interface element colouring satisfies the distinguishability criterion for those pairs of items specified by the developer. It has a worst-case value of 0 when all of the pairs of items that are required to be distinguishable are the same colour (i.e. occur at the same position in the abstract colour space). As the pairs of items become more distinguishable, D' increases, and the closer it is to 1 the better.

As will be discussed in section 4.5.7, D' is used to derive the final distinguishability score D for a molecule. For some molecules, $D' = D$, for other, a penalty is applied and D is less than D'. This changes the score, but not the overall semantics: both D' and D indicate how well a molecule satisfies the distinguishability constraints.

4.5.6 The readability contribution

It is clear from the literature reviewed in section 2.9.8 that luminance (light-dark) contrast is significantly more important for readability than an arbitrary colour space separation. Therefore, the evaluation of readability uses only luminance contrast to estimate the readability of the colours given to an item of text and its background.

The experiments by Zuffi et al. 2006; 2007 indicate that a difference in CIE lightness (ΔL) of 25–30 is sufficient for readability, irrespective of polarity. This gives a lower limit to the vertical separation in colour space between a textual interface element and its background. To ensure the text is not just readable, but easily readable, the target separation was increased to a ΔL of 48, based on the results of some informal experiments. This amount of contrast conforms to the published guidelines and also provides some safety margin in excess of Zuffi's minimum difference guideline in case the user's display is non-optimal. The heightened contrast could also be helpful to users with some vision impairment.

The readability evaluation function takes into account the luminance difference in a textual item from its background, but does not include the polarity[11] of the difference in its evaluation of the readability. During a later colour scheme personalisation phase (sec. 4.8), the user may swap the lightness values of all the colours in a colour scheme, causing light colours to become dark and vice versa, inverting the polarity of all textual contrasts. As will be discussed in section 4.8.3, this lightness inversion is a highly desirable design option, but as it makes it impossible to rely on the polarity of any item of text, polarity is not considered when assessing readability.

[11] Positive polarity text: dark text on a light background, with negative being the other way around (sec. 2.9.8)

The function used to assess readability for a pair of colour atoms is:

$$\begin{aligned} s_{i,j} &= N(l_{i,j}) \\ r'_{i,j} &= 1 - s_{i,j} \end{aligned} \tag{4.17}$$

where:

$l_{i,j}$	–	difference in lightness between item i & j in abstract colour space
$s_{i,j}$	–	similarity in lightness between item i & item j.
		If $(s = 1)$ they are most similar (i.e. there is no lightness difference)
$r'_{i,j}$	–	readability score of item i against item j; higher is more readable: $0 \leq r' < 1$
$N(\cdot)$	–	the normalisation function – equation 4.3

The overall raw readability score for a colour molecule is given by:

$$R' = \frac{1}{k}\sum_{i=1}^{n}\sum_{j=i}^{n} r'_{i,j} \quad \forall i,j \text{ where } r_{i,j} = \text{true} \tag{4.18}$$

where:

k	–	the number items with readability constraints
n	–	the number of colourable items in the interface
$r'_{i,j}$	–	readability score of item i against item j (equation 4.17)
$r_{i,j}$	–	the Boolean readability constraint predicate[12]defined on page 120

4.5.7 Ensuring that critical colouring flaws are not overlooked

When combined, the four contributing terms allow evaluation of a colour scheme in a manner that should correlate with the appeal of a colour scheme to an average human observer, but there are circumstances in which it will fail.

Minor reductions in the colour strength balance and wireframe alignment contributions to a colour scheme's fitness will not significantly impact the quality of the colour scheme as a whole. However, the same argument does not apply to textual readability or distinguishability: if any item of text is unreadable, or any control is indistinguishable, the impact on the overall score may not be great, but the interface will be unusable[13]. Therefore, if any item of text is unreadable, or any pair of items that should be distinguishable have colours that are not sufficiently different,

[12] The readability constraint returns true if item i & j must be sufficiently distinct to be easily readable.

[13] There are exceptions: not all broken distinguishability constraints will render an interface unusable. The developer may have defined some distinguishability constraints for aesthetic reasons, and if these are not enforced, appearance, but not usability, will be affected. However, as there is no indication of the reason for any constraint, all distinguishability constraints are enforced.

the overall readability or distinguishability score for the colour molecule is multiplied by a penalty factor. The value of this multiplier is progressively reduced from one to its minimum value over a range of readability or distinguishability, to model the fact that text becomes progressively less readable (or distinguishable) over a small range of colour differences.

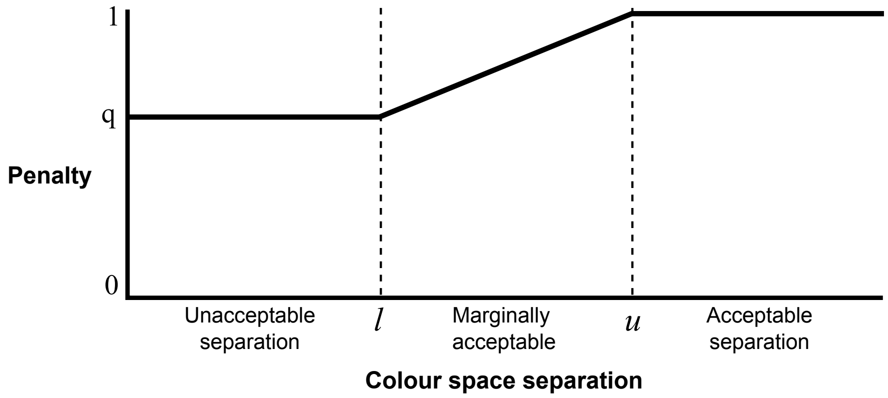


Figure 4.10: The penalty function applied to the overall scores for both distinguishability and readability to ensure all items are easily readable or distinguishable

Although the threshold values differ, the same method is used to calculate the penalty (if any) to be applied to the distinguishability scores (for user-defined pairs of items) and readability scores (for textual items against their backgrounds). The method of calculating the penalty factor will be detailed using readability as the example.

The raw score (R' defined above) is not penalised if all the readability constraints are satisfied: that is, if the lightness differences between all items and their backgrounds – when translated from the abstract colour space to real colours – are sufficient for good readability. If the readability of any of the item pairs is unacceptable, the raw score is scaled down to a fraction of its initial value, significantly penalising the fitness score for the molecule. For the intermediate cases, where a text item is discernable but not yet sufficiently different from its background to be easily readable, the raw score is a linear interpolation between penalised and raw values.

The readability score for each item-background pair is translated to an acceptability rating by a function that returns 0 if the readability score (v) is less than a lower threshold (l), returns 1 if v is above an upper threshold (u), or returns an interpolated value if $l < v < u$. Figure 4.10 shows the form of this function. The lower (l) and upper (u) thresholds indicate the thresholds of marginal and acceptable readability.

The rating of acceptability R_a is given by:

$$R_a(v,l,u) = \begin{cases} 0 & \text{if } v < l & \text{— contrast is unacceptable} \\ 1 & \text{if } v \geq u & \text{— contrast is acceptable} \\ (v-l)/(u-l) & \text{if } l \leq v < u & \text{— item is discernable but not acceptable} \end{cases} \tag{4.19}$$

The penalty function used in the prototype Colour Harmoniser is shown in equation

4.20. This function finds the acceptability rating for all pairs of items that should be readable or distinguishable. All of these values will be between 0 and 1. As it is desired for any single item being unacceptable (0) to affect all values, the values are multiplied together (the multiplicative product $\prod_{i=1}^{k} R_a(v_i, l, u)$ in the equation), so that:

- any single unacceptable readability or distinguishability value will force a 0 result.
- if items are acceptably readable or distinguishable, the product is 1.
- if all items are discernable but not yet all acceptable, the result is between 0 and 1.

This multiplicative product is then combined with the fully penalised value (q), so that: if the multiplicative product returns 0, the original readability or distinguishability score is scaled down by q (i.e. qx); if the product is 1, the original score x is returned unchanged; and otherwise the result is an interpolation between qx and x:

$$P(x) = x \left(q + (1-q) \prod_{i=1}^{k} R_a(L_{i_v}, l, u) \right) \tag{4.20}$$

where:

$P(x)$	–	returns the final value of x , the raw readability/distinguishability score
x	–	R' or D'— the molecule's overall Readability/Distinguishability score
$R_a(\cdot)$	–	the rating of acceptability interpolation function – equation 4.19
q	–	$0 < q \leq 1$ – the penalty factor for unacceptable scores (e.g. 0.8)
L_m	–	list of Readability/Distinguishability constraint pairs for the molecule
k	–	the number of item pairs with a Readability/Distinguishability constraint
v_i	–	the Readability/Distinguishabilty score of a specific item/background pair
l	–	discernability threshold value for Readability/Distinguishability
u	–	acceptability threshold value for Readability/Distinguishability

Using the threshold values for discernability and acceptability derived from human perception, this function modifies the raw score for the readability (or distinguishability) term of a molecule. It is not intended as an exact analog of a human evaluation, but one in which the raw score is unchanged if all required elements are readable (or distinguishable), but a single unacceptable element penalises the overall score. The penalty is reduced once all elements are at least discernable, and removed completely when all are acceptable.

The threshold values used in the Colour Harmoniser prototype are 10 & 20 ΔE for distinguishability (were determined empirically, and are discussed further below), and the 24 & 48 ΔL[14] for readability. The penalty factor q is 0.8, which penalises the raw

[14] The rationale is discussed in section 4.5.6.

scores by 20% if any constraint is unacceptable. The 0.8 is arbitrary, but sufficient to ensure that flawed schemes are significantly downgraded.

The ΔE limits for calculating the separation required for visually distinguishable elements are based on the assumption that the size of the elements (e.g. buttons, panels, and backgrounds) is substantial. However, if the target interface includes semantically important finely detailed graphical elements (e.g. the thin lines and symbols used in a CAD package), the distinguishability term would need to use the lightness difference (ΔL) in its evaluation, not the colour space difference (ΔE) (Klassen et al., 1998).

4.5.8 The fitness function

The final fitness function is made up of four scores, each assessing a component important to the use of colour in a user interface. Each of the four terms has a uniform range, which ensures that no one term is so heavily weighted that it swamps contributions from others. The four terms are combined as a weighted sum:

$$F = w_i C + w_j W + w_k D + w_l R \tag{4.21}$$

where:

F	–	the overall fitness function score	
C	–	the colour strength balance score	equation 4.5, p129
W	–	the wireframe alignment score	equation 4.13, p130
D	–	$P(D')$, final distinguishability score	equation 4.16, p131
R	–	$P(R')$, final readability score	equation 4.18, p133
P	–	the penalty function	equation 4.20, p135
w_i, w_j, w_k, w_l	–	the associated weights	

The fitness function – summary

A fitness function that can be used to assess an interface colour scheme on both aesthetic and usability criteria has been described. Four terms are combined as a weighted sum. The first two of the four terms evaluate factors relating to aesthetics, and the remaining two evaluate factors relating to usability.

The first aesthetic term evaluates the overall colour strength balance: whether or not the colours used in an interface colour scheme have saturations and lightnesses that are appropriately related to their size. Balanced colour (complementary, split-complementary and elliptical) schemes are assessed for their balance around mid-grey, and unbalanced (offset) schemes (monochromatic and analogous) are evaluated using the inverse-area rule. The second aesthetic term evaluated is wireframe alignment, which evaluates how closely the interface element colours are to those of the preferred colour scheme, as defined by a path in the colour space. To allow black, white and grey elements to be used in the colour scheme (as in common with interfaces), the wireframe can include the lightness axis, in which case the colour balance evaluation is modified so achromatic areas do not affect the colour strength balance.

The two usability-related terms evaluate whether textual interface items are coloured so that they are easily readable, and whether the pair of items that the developer has indicated should be distinguishable satisfy this requirement. To ensure that schemes that are unusable (because of unreadable text or indistinguishable controls) have a score that reflects this impaired usability, a penalty may be applied to the overall readability or distinguishability score if one or more elements are not clearly readable or distinguishable.

The fitness function outlined in the preceding sections does appear to score colour schemes as desired. Those with poor readability or indistinguishable elements are penalised, and those with highly rated colour balance and wireframe alignment are frequently acceptable. Exactly how well the fitness function scores correlate with those of human evaluators will be examined in detail in section 5.2.

4.6 Colour molecule optimisation

The fitness function allows any placement of colour atoms on or off the wireframe to be evaluated for its suitability as an interface colour scheme. Positioning atoms at evenly spaced intervals (as described in section 4.4) enables many different colour schemes to be created, but the chance of discovering an ordering that satisfies all of the constraints with the atoms in their initial predefined positions is exceedingly small. However, using the fitness function to evaluate the schemes, it is possible to determine which of the schemes score more highly, and adjust the positions of the atoms to increase the molecule's fitness score, and therefore its suitability for use as an interface colour scheme.

The arrangement of atoms that maximises all four fitness function parameters simultaneously cannot be found deterministically. Even if it could be, the result would be a single theoretically optimal colour scheme. The models and the heuristics used to create the schemes are based on the colour relationships that have been found to underly colour schemes with wide appeal. However, the models only specify the interrelationships between the elements, not their hues. So, to render schemes derived from such a model, it is necessary to choose a hue (a wireframe rotation) for each scheme. For high scoring schemes, this may immediately result in an acceptable colour schemes, but if not, it is hoped that simple adjustments will allow their transformation into schemes that do appeal.

Therefore, in order to satisfy the subjective element in human aesthetic colour scheme preference, the viewer must be included in the optimisation process. In the evolutionary approach used by Kelly (1999), the designer acts as the "evaluation function" and selects preferred sets of colours from those generated by the computer. This avoids the need for a computational objective function, but as noted, due to the size of the search space, it can sometimes be difficult to find good initial candidates: from a human perspective the process is tedious and slow. It may also be difficult for a person to choose the better alternative from a pair of colour schemes they consider unpleasant. This difficulty is circumvented in two ways in the Colour Harmoniser architecture: the creation of schemes based on wireframes, and mechanical optimisation of the schemes before they are presented for assessment.

However, before users can tailor a scheme to their liking, it must satisfy all the constraints: that is, it must have high fitness scores for all four criteria. Only then, in the final stages, is the human element included. The developer is presented with a preselected set of colour schemes, all of which have high fitness. They can then select a preferred scheme from those offered and perform the final optimisation manually.

4.6.1 Optimisation requirements

The desired result of the optimisation phase is a small set of optimised colour schemes, all of which have high overall fitness. As it is impractical to evaluate all possible arrangements of the colour atoms on a wireframe, an optimisation process is used to select high scoring schemes.

The requirements of the optimiser are these:

1. it must be able to locate multiple solutions approaching the global maxima.
2. it must be able to escape from local maxima.
3. it must function with an evaluation function that is both non-linear and discontinuous.

There are many commonly used optimisation techniques, none of which is the best for all possible problems (the so-called "No Free Lunch" theorem (Wolpert and Macready, 1997)). Given that the colour scheme evaluation function is discontinuous, conventional gradient descent algorithms (Snyman, 2005) cannot be used. The choice is therefore a stochastic algorithm or a relaxation algorithm (e.g. Ryall and Marks (1996)). The former is preferred because relaxation algorithms are poor at global search. Of the stochastic algorithms, the most commonly used is the genetic algorithm; other possibilities include simulated annealing and stochastic hill climbing.

Stochastic methods are computationally expensive, but do not use gradients and can therefore function with a discontinuous fitness function. They perform a wide-ranging search of the solution space, and are less constrained by the initial starting configuration than gradient-descent and relaxation optimisers. However, the solutions found may not be near a global maxima, and, on occasion, the optimiser may converge prematurely, or may not converge at all.

With appropriate modifications, it would be possible to use any of previously mentioned optimisation techniques. However, the genetic optimiser requires no modification to work with a non-linear and discontinuous fitness function, performs a wide-ranging search, and is capable of preserving multiple good solutions within a population. It was therefore chosen as the optimisation method to increase the fitness score of a candidate set of colour molecules.

4.6.2 Genetic optimisation applied to colour molecules

Genetic optimisers function by maintaining a population of candidate solutions, each of which is composed of several distinct sub-factors (Marsland, 2009). Members of this population are evaluated and ranked and aspects of the better scoring candidates are combined and used as the basis for the population of the next generation.
The requirements for the use of a genetic optimiser are (Marsland, 2009):

- *an objective function*: a means of evaluating the overall merit of any potential solution. The fitness function has already been described can be used as an objective function for the evaluation of interface colour schemes;
- *a string representation*: a method of representing potential solutions (candidates) as a string. In the proposed architecture, the candidates are colour molecules, and the sequence of colour atom coordinates within each molecule is the string

representation of any potential solution. An abstract colour scheme is the set of the (x, y, z) positions of n colour atoms for any given colour molecule.

$$m = a_i(x_i, y_i, z_i) \quad \text{for } i = 1 \ldots n \tag{4.22}$$

where:

m	–	a colour molecule (an abstract colour scheme)
a_i	–	the colour atoms of molecule m in abstract colour space
n	–	the number of colour atoms (independently colourable areas)

- *genetic operators*: methods of modifying member of the existing population in order to derive new members. The two most commonly used genetic operators, crossover and mutation, are inspired by biology.

The overall genetic optimisation process can be outlined as follows:

1. *create an initial random population:* this provides many wide-ranging samples of the solution space.
2. *select the parents for the next generation:* evaluate the fitness of each member of the population and assign those with higher fitness a greater chance of selection.
3. *breeding new solutions using crossover:* perform a global search by creating new members of the population by exchanging (crossing over) characteristics between two members selected as parents.
4. *mutate:* slightly perturb the genotype of some of the population. This enables local variations to be examined (a local search).
5. *test for continuation:* stop if the best solution is satisfactory. Otherwise, go back to step 2 (using the newly modified population).

The mutation and crossover operators applied to colour molecules

Mutation – random change in the genotype – can be implemented by perturbing the position of a randomly selected colour atom within the population. The crossover operator can be implemented by slicing the string representation of a colour scheme between any two atom positions and exchanging the remaining positions after the slice point. For example, given two colour schemes represented by colour molecules M_a and M_b, two derivative schemes can be created by crossing over part of one scheme with part of the other:
If two colour molecules prior to crossover have the arrangement of atoms shown in the table below:

Arrangement	Item:	1	2	3	4	5	6	7	8	9
M_a		a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
M_b		b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9

then after crossover, having exchanging items 1 to 3, the new arrangement will be:

Arrangement	Item:	1	2	3	4	5	6	7	8	9
N_a		b_1	b_2	b_3	a_4	a_5	a_6	a_7	a_8	a_9
N_b		a_1	a_2	a_3	b_4	b_5	b_6	b_7	b_8	b_9

The two schemes resulting from the crossover are not necessarily better than, or even as good as, the old. To avoid degrading the population with schemes that are worse after the crossover/mutation than before, the fitness of all four schemes in a crossover (the two parent schemes and the two derived schemes) can be evaluated and the highest scoring pair propagated to the next generation. This technique (tournament selection) will result in faster convergence by not automatically replacing the parents with the new schemes if the newer schemes are worse, but it does reduce the diversity in the population and may result in premature convergence. Elitism – the automatic copying of a small number of the best scoring solutions into the next generation – is another improvement that ensures the best scoring schemes are not lost during the creation of the next population.

The idea underlying a genetic optimiser is that the better solutions are combined to produce the next generation. This can be achieved by making the probability of a scheme being used as a parent proportional to its relative fitness (i.e. its fitness as a fraction of the total population fitness). Best scoring schemes would have the most offspring, lesser scoring schemes proportionally fewer, which helps ensure that attributes from good colour schemes will be preserved and those from lesser schemes will be dropped from the population.

The crossover mechanism introduces some random variation, as the parents of the next generation are chosen randomly. However, if only crossover was used to produce succeeding generations, the population would converge to the scheme with the highest fitness score, but only using the originally chosen positions of colour atoms on the wireframe, which are a very small subset of the possible wireframe positions. The introduction of further randomness will enable a more wide-ranging search. Two other mechanisms can be used. The first is to include some colour molecules in the initial population whose atoms are not on the wireframe, but are positioned randomly. The fitness of these random schemes is unlikely to be good, but during the crossover phase, the blending of atom positions from molecules with moderate fitness scores should result in a more extensive search of the colour scheme solution space. The second method of introducing random variation is mutation, where the position of a randomly selected colour atom is perturbed slightly. This moves the atom from its current position (possibly off the wireframe), which enables the small changes in colour to be tested.

These two techniques enable both global and local optimisation to occur: crossover and the introduction of randomly positioned molecules can induce large changes which sample completely different sections of the colour solution space, and the mutations explore nearby solutions.

4.6.3 Multiple solutions from the optimisation

Optimising towards a single scheme is undesirable in the current context as, even if all the constraints are satisfied and the scheme conforms to the widely accepted guidelines for colour harmony, the colours used may not suit the designer's personal taste. While the top scoring members of a single population will differ, the differences can be small. Therefore, rather than selecting multiple solutions from a single population, a small number of separate populations (ten) are optimised independently in the Colour Harmoniser prototype. The different initial populations will have differing wireframe orderings and different colour atoms in the randomly coloured molecules. These sources of variability, when combined with random mutation, result in the ten populations being visibly different at the end of an optimisation. This process yields multiple optimised colour schemes for the user's consideration.

The result of the optimisation phase is therefore not one, but ten schemes, all designed using the same user-defined constraints and groupings, and all with the same underlying wireframe. Because of differing initial conditions, the ten colour schemes are unlikely to be the same, but will be visually similar, as the solutions are constrained to use colours either on, or close to, the wireframe. As the colours are all from the same orientation of the wireframe[15], the same hues will be used to colour the user interface elements in different ways.

This is similar to the SmartColor system (Wakita and Shimamura, 2005), which displays the best and several less optimal options (see sec. 2.12), but in the Colour Harmoniser architecture, the multiple colourings are intended to address the differences in the personal tastes of the designer and end-use suitability, not (as in the SmartColor system), the incomplete specification of constraints.

Given the very large number of potential solutions, it would be possible to offer the user many more than ten potential solutions. However, it is not necessarily beneficial, as user satisfaction is higher when the pool of options from which the choices is made has a small number of candidates (Schwartz, 2005). This seems related to the users' concern about missing a good option being higher than their satisfaction with their current choice. If the number of initial options is limited, but designed to exceed some arbitrary minimum quality, users can be fairly confident that they haven't missed a good option, and are therefore happier selecting one option from a small number of choices than from many.

4.6.4 Avoiding premature convergence

The genetic optimisation technique allows the search to range widely within the solution space and will frequently result in good solutions. However, it has its own set of

[15] Note that the tweaking of the real colour scheme that follows the optimisation allows the user to rotate the wireframe and so change the hues, so this restriction is only temporary.

disadvantages. For example, the derivation of solutions is opaque. While the logic underlying the genetic optimiser is well-defined, the system operates as a "black box", with no explanatory power regarding the derivation of any single solution.

A more significant problem relates to the stochastic approach – there is no guarantee of convergence to a good solution. This is partially addressed by selecting parents in proportion to their fitness. However, if there is insufficient randomness left in the current population, the randomness induced by mutation may be insufficient to allow the optimiser to escape from a poor solution. This is known as premature convergence, and can be addressed by introducing more variability into the population. Two methods are commonly used. The first is to forcibly introduce significant randomness into the population and restart the optimisation, which introduces diversity while preserving some of the good solutions. The second is introduce some diversity into the population by introducing good candidates from another population. An example of this technique is that of "island crossover" or "niching" Marsland (2009). The name "island crossover" is an analogy for the situation in which two populations of a biological species have evolved separately on different islands, but occasionally, some members of one population travel to the other and introduce some genetic diversity. However, introducing a strong member of one population into a weaker and genetically less diverse second population introduces the risk of both populations converging to the same (or substantially similar) solutions. As the express reason for maintaining multiple separate populations is to generate different colour schemes – to maintain diversity – island crossover was not used. Instead, to enable the optimiser to recover when progress had stalled, the dynamic restart algorithm (Solano and Joyner, 2007) was used. This monitors the change in fitness of the best scoring scheme in each population and, if no progress has been made for a preset number of generations, atoms within the population are randomly coloured and the optimisation restarted. This significantly reduced the number of stalled optimisations.

In the prototype implementation of the Colour Harmoniser, the string representation of a colour molecule used by the optimiser is $(x_1, y_1, z_1 \ldots x_n, y_n, z_n)$, where each (x, y, z) is the real[16] position of a colour atom in abstract colour space. There were ten independently optimised populations, each of which had 100 members, 80 of the each population had the colour atom on the wireframe in various orderings, and in the remaining 20, the colour atoms were positioned randomly. The crossover points were randomly selected between atoms and therefore only crossover complete atom positions. The mutation rate was 5%: that is, each molecule in a generation has 5% chance of being mutated[17]. Both elitism – copying the best four solutions – and tournament selection were used when creating the next generation. Although the number of iterations required for the optimisation to converge varies depending on the number of colour atoms, the number of user-defined constraints and the wireframe shape, typically 100 to 150 iterations is sufficient.

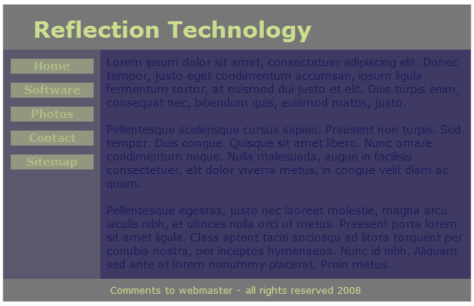
As an illustration of the variability within a population, four schemes from a single

[16] therefore this is a real-valued genetic optimiser, not one that (directly) uses bit-strings.

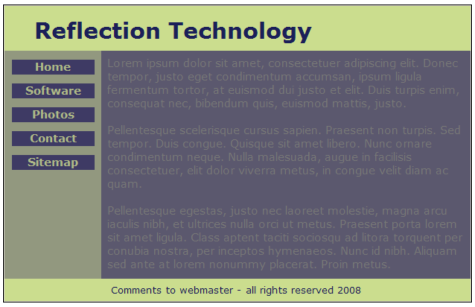
[17] with the resulting atom position being scaled if necessary to lie inside the unit sphere in the abstract colour space.

unoptimised population are shown in figure 4.11. As can be seen, in the top ranking scheme, the colours are pleasant, but the readability is poor. The best random scheme has good readability and all the elements are clearly distinguishable, but the lack of coherence in the scheme results in a poor wireframe alignment score (0.17) and the garish use of colours results in a poor colour balance score (0.19). After optimisation, the best scoring schemes are free of these defects, as can be seen in the two top scoring schemes shown the lower two schemes in figure 4.11.

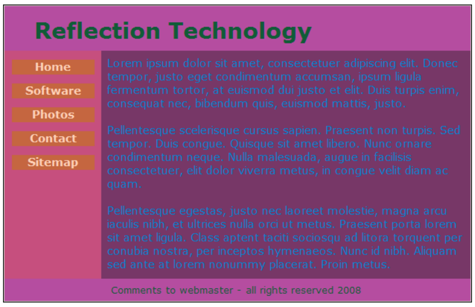
The variability in the resulting ten optimised schemes can be seen in figure 4.12, which clearly shows how, even with colours constrained to a wireframe, significant differences in overall appearance can result.



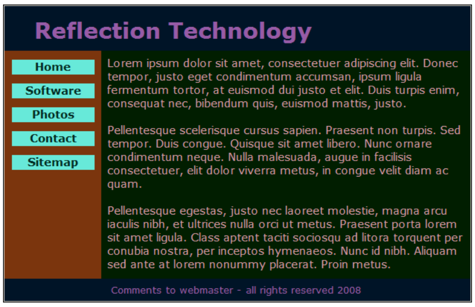
The lowest scoring molecule with atoms on the wireframe from an unoptimised population (score of 2.38).



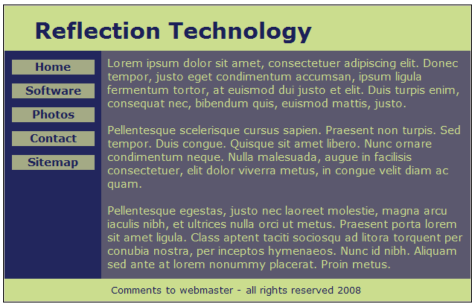
The highest scoring molecule with atoms on the wireframe from an unoptimised population (score of 3.28).



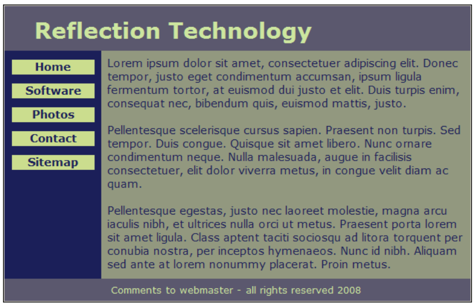
The lowest scoring of the random arrangements from an unoptimised population. The scheme has poor contrast and little coherency in the underlying colouring (score of 1.68).



The highest scoring random scheme from an unoptimised population. This has better contrast (score of 2.2).



After optimisation, the top scoring scheme from one of the ten populations (score = 3.76).



After optimisation, the top scoring scheme from another of the ten populations (score = 3.70.)

Figure 4.11: The top four images show colour schemes from a single population before optimisation: some molecules have the atoms spaced evenly on the wireframe (top row), and some have the colour atoms positioned randomly in the colour space (middle row). As can be seen the schemes have different mixes of readability, distinguishable elements, and pleasing colour schemes, and some are more usable than others. This is reflected in their fitness function scores (out of 4). After optimisation, as shown by the schemes in the bottom row, readability has improved, all the items are distinct, and the saturation of colours is appropriate to their area.

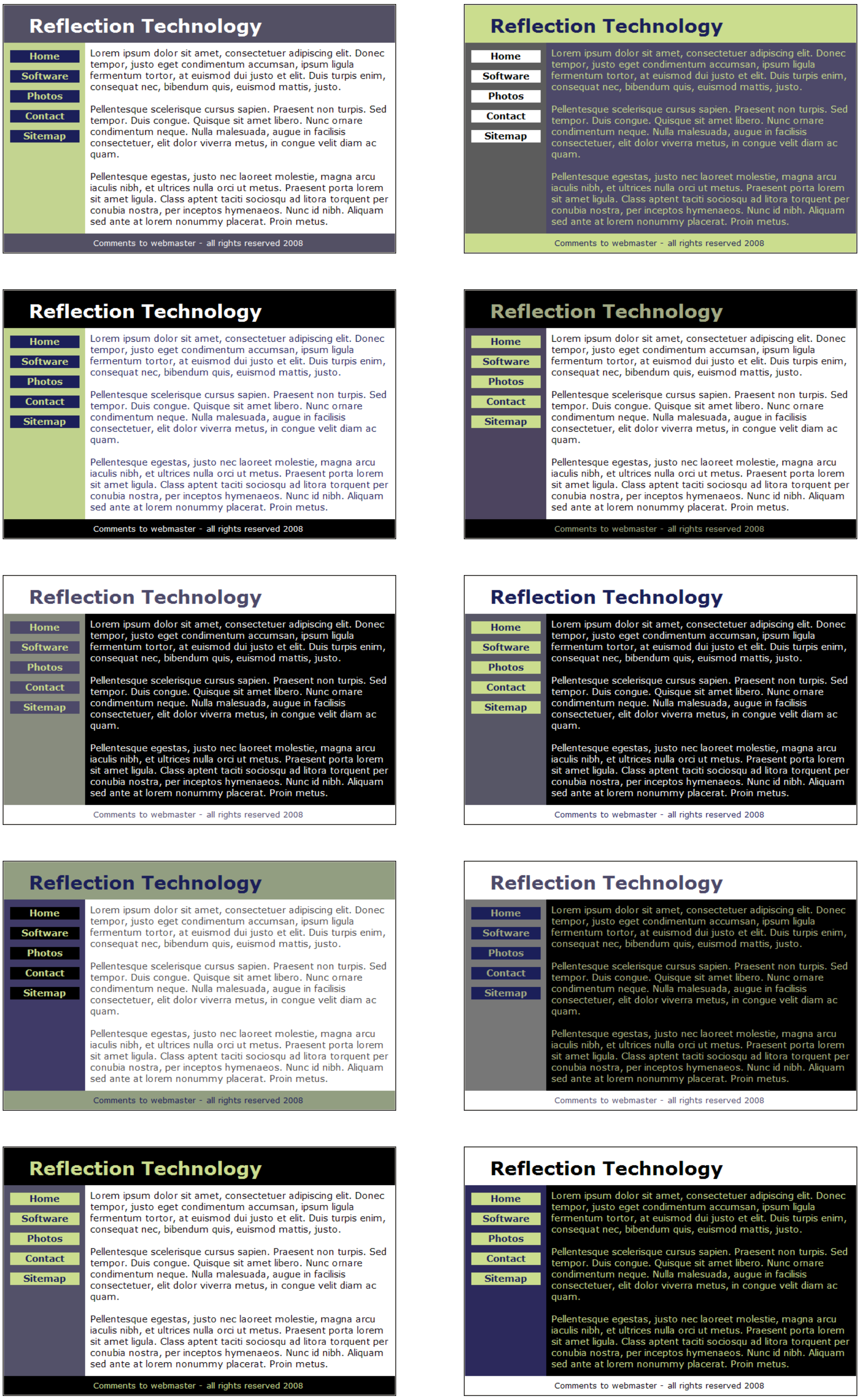


Figure 4.12: The ten colour schemes resulting from optimising ten different populations, all using hues defined by the current wireframe rotation angle. The images illustrate the diversity of schemes possible using the same underlying colours. The basis for all of these colour schemes is a split-complementary wireframe augmented to allow the positioning of items on the black-white axis. All would be suitable for online use.

4.7 Real colour schemes from an abstract molecule

The abstract colour space is assumed to be perceptually uniform and spherical, as its shape is defined by the sweep of the wireframe extremities. Any colour scheme designed in this abstract colour space may, because of the perceptual uniformity, be subsequently reoriented without invalidating any of the design criteria used by the optimiser to determine the relationships between the atom colours.

Transforming the abstract colour coordinates of the atoms from the abstract colour space into a real perceptually uniform space (e.g. CIELAB) would allow an arrangement of atoms in the abstract space to be mapped to RGB colours for display. To alter the scheme colouring, the molecule could be reoriented in the abstract colour space prior to mapping into CIELAB, or mapped to CIELAB and then reoriented: the effect is the same. Both methods would preserve perceptual relationships between the colours of the atoms, and allow the transformation of abstract coordinates to displayable colours.

However, the spherical abstract colour space is an idealisation; real perceptually uniform spaces are not spherical. This could be handled in two ways; by allowing the shape of the perceptually uniform space to influence the design of the colour scheme, or designing the colour schemes in an ideal abstract colour space, and handling any mismatches related to available colours in a post-design phase:

allow the shape of the colour space to influence the design:
The first option – allowing the shape of the perceptually uniform space to influence the design of the colour scheme – could be implemented by designing the colour scheme directly in the irregularly shaped perceptually uniform space or equivalently, by an abstract colour space that is a scaled version of a real perceptually uniform colour space. This ties the design of colour schemes to the characteristics and shape of the perceptually uniform space. The orientation of the wireframe affects the design of the colour scheme, as the possible lengths of the wireframe segments will be limited to the limits of the space and, as can be seen in figure 4.13, this varies significantly at different rotational angles.

design the colour schemes in an idealised space:
Rather than allowing the odd shape of perceptually uniform space to complicate the design, the colour scheme could be designed in an ideal spherical abstract colour space, and the mismatch between the shapes of the abstract colour space and a real perceptually uniform space handled at a later stage.

This separates the design of colour schemes from the constraints imposed by the irregularity of the shapes of real perceptually uniform spaces, and more importantly, enables the design of the colour scheme to be independent of a particular wireframe orientation in the perceptually uniform space. This is advantageous, as the wireframe segment lengths would vary depending on the its orientation within the perceptually uniform space, which would mean that the positioning of items within the scheme could vary, depending on the wireframe orientation.

One of the central ideas being explored in this research is the viability of designing colour schemes in terms of the relationships between the items being coloured (but not their hue), to allow the hue to be freely chosen at a later stage. Therefore, the second option was the one chosen for the Colour Harmoniser prototype: colour schemes are designed in an ideal spherical abstract colour space without considering hue or any particular the wireframe orientation.

This approach leads to three colour spaces being used during the design of a colour scheme:

- an abstract colour space that is assumed to be perceptually uniform and spherical in shape (due to the sweep of the wireframe extremities).
- an intermediate, perceptually uniform, colour space into which the abstract colour schemes are mapped. The shape of this colour space will be limited by the gamut of the display device and it will have a highly irregular shape.
- an RGB (perceptually non-uniform) display colour space, the characteristics of which constrain the shape and size of the intermediate colour space. Coordinates in the intermediate colour space are mapped to the RGB space to find the actual colour of interface items.

The intermediate colour space could be any perceptually uniform colour space (CIELAB was used for reasons explained shortly), but it must be known before colour schemes can be optimised, as it can affect whether or not the resulting schemes will have readable text or distinguishable elements. The minimum required separations between items are specified in intermediate colour space (i.e. in CIELAB: ΔL for readability; ΔE for distinguishability). However, the optimisation (including determining whether or not the elements are sufficiently separated) is carried out in the abstract colour space. Therefore, the equivalent distances (to $\Delta L/\Delta E$)in the abstract colour space must be known prior to optimisation, so the optimiser can use these values ensure that the colours in a scheme, when eventually transformed from abstract coordinates to real colours, will satisfy the readability and distinguishability constraints.

Consequently, before abstract colour schemes can be optimised, the inverse mapping function (from intermediate colour space coordinates to abstract colour space coordinates) must be available, so that the required separations in abstract colour space can be calculated. If this transformation is not available, or if the intermediate or final RGB colour spaces are unknown, the design of colour schemes in the abstract space cannot guarantee to produce satisfactory real colour schemes.

The intermediate colour space used in the Colour Harmoniser prototype was the CIELAB colour space, because it is approximately perceptually uniform, mathematically tractable, and widely used. To display a colour specified in CIELAB coordinates, these coordinates must be mapped to an RGB colour space and, as discussed in section 2.6, the current standard for non-professional computer displays is sRGB rendering. The Colour Harmoniser prototype uses sRGB as its display colour space. The use of this standard means that colour schemes created in the abstract colour space and then mapped to CIELAB and then sRGB will, when displayed on uncalibrated displays, look no different (on average) from images created by software engineers and web site developers using equipment based on the same sRGB standard.

4.7.1 Mapping abstract colour space onto a perceptually uniform space

At the end of the optimisation phase, a small number of colour schemes (ten in the prototype implementation) have been optimised. Each of these schemes is in the form of a colour molecule whose colour atom coordinates correspond to "colours" in an abstract colour space. To convert these to real colours, the abstract coordinates must be mapped into a real colour space. Differing scalings, translations and rotations will result in related, but differently coloured, schemes. Translation does not appear to serve any useful purpose, nor does it appear to be necessary, so it is not considered further.

Differing scalings and rotations will yield sets of different real colours for each of the colour atoms (and the associated items in the target interface). For a mapping to preserve the characteristics of the optimised colour scheme, there are two constraints. Firstly, if the colour atom positions are to be moved after the mapping, the colour space within which the movement occurs must be perceptually uniform. Secondly, the y axis in the abstract colour space must be mapped to the lightness (or value) axis of the perceptually uniform colour space. This is assumed during the optimisation and used to ensure sufficient lightness difference to provide easily distinguishable text. The (x, z) axes in the abstract colour space are therefore mapped to the saturation plane of the perceptually uniform colour space.

Once the coordinates have been mapped from the abstract colour space to a perceptually uniform colour space, they can then be mapped to sRGB for display. However, there are two significant complications: the gamut of the sRGB space is smaller than that of the CIELAB space, and the shapes of the spherical abstract colour space[18] differ significantly. As can be seen from figure 4.13 which shows the sRGB gamut projected into CIELAB coordinates, the shape of the perceptually uniform space is anything but spherical. This shape mismatch is not specific to the CIELAB space, but is a characteristic of human visual perception, and for the gamut shown, the limits of the sRGB display technology. The Munsell colour space and perceptually uniform approximations (such as the CIELUV colour space) have similar highly irregular shapes, and would cause similar difficulties. Nevertheless, for the system to function, a way must be found to map the coordinates of colour atoms from the abstract colour space into the CIELAB space.

If it is acceptable to limit the saturation of colours used in the final colour scheme, then a simple linear scaling of the abstract coordinate values could be used, with the scaling factor chosen so that all the colours lie within the gamut of the display device. However, because the minimum diameter of the gamut-limited perceptual space is much

[18] it is the volume swept out by the extremities of a wireframe with a maximum radius of 1.

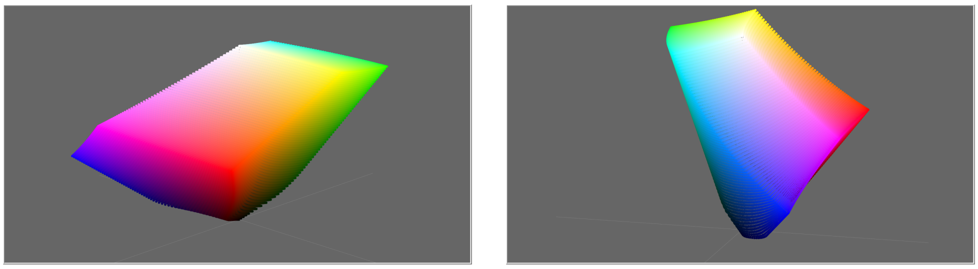


Figure 4.13: Two views of the sRGB gamut in the perceptually uniform CIELAB space. As can be seen, there is a significant mismatch between the shapes of the spherical abstract colour space and the shape of sRGB projection.

less than its maximum diameter, the colour range of the resultant schemes would be severely restricted with all the colours being quite desaturated. To keep all possible tilts[19] and rotations of the wireframe, only colours within the dotted sphere shown in figure 4.14 could occur in the final colour scheme. This disallows a very wide range of colours and is unduly pessimistic.

The colours used by a particular colour scheme are not distributed uniformly within the sphere, but lie on a wireframe within the sphere. When the projected wireframe points align with the shape of the longer axis of the perceptually unform space, a larger scaling factor is possible while staying within gamut. However, there is no single scaling factor that will give the largest possible range of in-gamut colours at all wireframe rotations. If a low scaling factor is used to ensure all colours will be within gamut, then most of the colours will be very desaturated, but the use of any larger scaling factor will mean the projected abstract colour atom positions, for some wireframe orientations, will be out of gamut.

In attempting to satisfy the distinguishability or readability constraints and to maximise the contrast or colour difference between two coloured objects, the optimiser often positions items at the endpoints of the wireframe. If it is necessary to use highly saturated colours and to keep all the colours within gamut, the scaling factor would depend on the rotation angle of the molecule. However, one of the initial goals of the Colour Harmoniser approach was to produce a colour selection tool in which the user could freely alter the overall colours (by rotating the wireframe), and because of the properties of the perceptually uniform colour space, the relationships between colours would remain as intended. While the scaling factor could change as the wireframe rotates so the scaled wireframe uses as much of the irregularly shaped perceptually uniform space as possible, the differing scale factor would cause the saturation of the colours in the scheme to vary wildly, which is obviously undesirable.

During the optimisation, the scaling factor must be available to the optimiser so that it can arrange for the separations between items[20] in the abstract colour space to

[19] While the wireframe used has a 45° slope, this is not required. There is another degree of freedom not explored here, that of tilting the wireframe so that it becomes more or less vertical. Tilting the wireframe towards vertical will lighten and desaturate the colours. Tilting it towards horizontal after optimisation risks introducing readability difficulties.

[20] those required to be distinguishable or readable

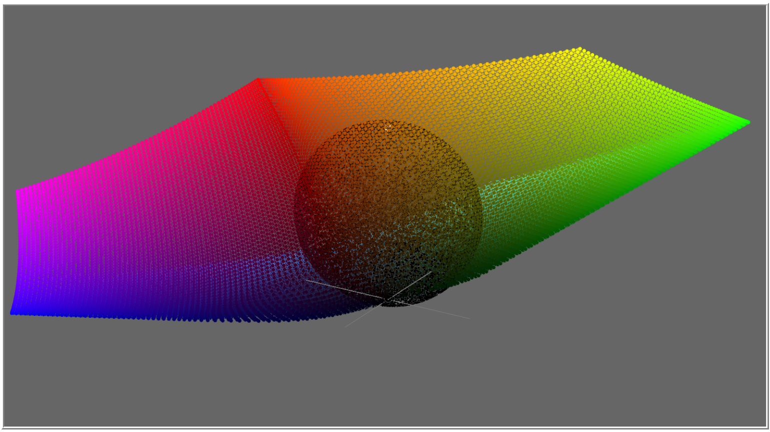


Figure 4.14: Only colours within the dotted sphere are within gamut at all possible orientations of the wireframe. If it was strictly required that all the colour transformed from the abstract colour space be "correct" at all orientations of the wireframe, only those colours within the sphere (most of which are quite desaturated) could occur in the resulting colour schemes. Sacrificing the wide range of colours outside the sphere is too great a penalty for the use of a small fixed scaling factor as an abstract-to-real transformation method to be seriously considered.

be such that, when the scheme is later mapped to real colours, the real colour differences will be sufficient to ensure that the constraints are satisfied.

4.7.2 The LAB scaling factor

As long as out-of-gamut colours are allowed for, mapping abstract coordinates to CIELAB may be accomplished by simple scaling. The y abstract coordinate axis ($-1 \rightarrow +1$) is mapped to the CIELAB $L*$ lightness axis ($0 \rightarrow 100$, corresponding to black–white). The other two abstract axes x, z mapped with a variable scaling factor (the *LAB scaling factor*) to the $a*$ and $b*$ CIELAB axes. Varying the LAB scaling factor will alter the radial distance in the CIELAB colour space, effectively controlling the maximum saturation that may be attained from points at the extremes of the abstract colour space.

If the LAB scaling factor is sufficiently large, the scaled values of the atoms in the abstract colour space may transform to CIELAB values that, when converted to RGB for display, are outside the gamut of the display device. The maximum values of the scaling factor before this out-of-gamut condition occurs will depend on the orientation of the wireframe.

The following methods could be used to address this:

constrain the LAB scaling factor: the most simplest and most conservative course is to limit the scaling factor so the scaled abstract colour space coordinates are always within the available gamut in the intermediate (CIELAB) colour space. However, this sacrifices most of the saturated colours: all colours outside the dotted sphere in figure 4.14 would be unavailable, so this option is discarded.

dynamically adjust the scaling factor as the wireframe orientation changes: a scheme could be optimised with an LAB scaling factor relying on the wireframe being positioned so that it is aligned with the maximal diagonal of LAB space. If this were so, the optimiser would use the available colour space in its positioning of colour atoms. If the same molecular arrangement is maintained as the wireframe is rotated, and the scaling factor is progressively reduced to keep the final colours within gamut, the inter-atom differences could easily be reduced to the point where the readability or distinguishability constraints are no longer met. The alternative – to perform the optimisation at the most constrained wireframe angle and alter the scaling factor as more of the LAB space comes within gamut – would result in the saturation of colours changing in an inexplicable way as the user rotates the wireframe to alter the hue. Neither of these conforms to the GUI predictability guideline Shneiderman (2000) so this option is discarded also.

produce separate optimisations at incremental rotations: this creates a separately optimised colour scheme at incremental rotations of the wireframe (e.g. every 5 or 10°). Later, when the user rotates the wireframe to change the hue, the appropriate optimisation for that wireframe angle would be selected, the abstract to CIELAB to sRGB mapping would be performed, and the interface recoloured accordingly.

This would produce schemes that include the correct and most strongly saturated colours possible at all wireframe rotations. It is computationally expensive and time-consuming, but more importantly, makes the interface interaction unpredictable. There is no relationship between the optimisations at one rotational angle and the next. Therefore, it is possible that as the available CIELAB space changes as the axis of the wireframe rotates, the optimiser might opt for a completely different arrangement of atoms along the wireframe (e.g. to handle atoms being contracted so much they are no longer distinct). From the user's perspective, this would appear as a complete change in the colour arrangement from one rotational angle to the next, again violating the interface predictability guideline. This option is discarded.

use the gamut to limit the rotation of the wireframe: it is possible to restrict the rotation of the wireframe to just those angles in which the CIELAB values are within gamut for the current scaling factor. This would not mesh well with a direct manipulation interface, as it would produce hard limits, stopping the rotation of the wireframe should the current LAB scaling factor exceed the range of displayable colours. This discontinuity in a direct manipulation interface is poor HCI design due to its lack of transparency. Even more disconcerting from the user's perspective, the hard limit points would change depending on the value of the LAB scaling factor. This option was discarded.

use gamut compression in the abstract-to-real mapping: There are many possible ways to map out-of-gamut colours to an available gamut (MacDonald, 1993).

One method is to uniformly compress the range of all the colours to bring out-of-gamut colours within range. This has the effect of desaturating the colours. It can be used when the relationships between the colours are more important than the absolute colour, especially given the human ability to adapt to colour shifts (Jameson and Hurvich, 1989; Shepherd, 1999; Wesner and Shevell, 1992). The optimiser, while balancing the constraints, will have used the LAB scaling factor in its calculations to ensure the readability and distinctiveness constraints are satisfied. Any reduction in luminance or colour space difference could result in a scheme that no longer satisfies the design constraints, so changing all the interface element colours to handle out-of-gamut colours is also unsatisfactory, so this option is also discarded.

alter only out-of-gamut colours: An alternative is leave those colours that are within gamut unchanged, modifying only those out-of-gamut colours that would be otherwise unrenderable. This approach could alter the relationships between the colours, which sounds like a poor compromise, but in practice, it works well for two reasons. Firstly, the out-of-gamut colours, while not strictly correct according to the colour model used by the Colour Harmoniser, are being produced as a direct result of the user's interaction and therefore change gradually. If the colours look wrong, they will only do so as the user modifies the controls, so the effect is a direct response to a user action. Secondly, this method does not introduce any unexpected colour variation (or lack of colour variation). All colours continue to change smoothly as the user moves the controls, and, although some colours might differ from those calculated, in practice this is not evident. This is a viable option.

Dynamic colour updates based on visible area

As explained in the previous section, the displayed colours may not be exactly as calculated. This is one source of error. There is another: the colour balance is based on the relative areas of interface elements, but an approximation is used for textual elements, and some interfaces use resizable elements (e.g. a splitter bar that can vary the area ratio between two panels[21]). To avoid this source of error, the interface elements must have the same areas as when the interface was characterised. However, this would mean no dynamic updating of text and no resizing of interface elements, not even the use of scroll bars[22]. These are unrealistic requirements.

The areas of the interface elements could be reassessed in real time and the correct area values used to create an updated colour scheme. This approach was been adopted for two reasons, one pragmatic, the other relating to good practice for GUI design. It is desirable, from both a conceptual perspective and an implementation perspective, to have the characterisation of the interface, the optimisation, and the human selection

[21] With a *splitter bar*, the overall area doesn't change, but making one panel bigger makes the other smaller.

[22] Scroll-bars would allow (for example) the text in the window to change as a document is scrolled, violating the "static text" requirement.

of the final colouring as separate phases. While this goal could be sacrificed, there is another more important consideration. Creating the abstract colour scheme requires an optimisation that is time-consuming and capable of producing several outcomes. To reoptimise using the exact area values while the interface is in use would require that the optimisation be run sufficiently quickly to keep the colours updated. Even if this was possible, using the current areas of interface elements would introduce dynamic variations in their colours as the colour strength factor was reassessed.

From the user's perspective, the colour scheme would appear to change arbitrarily. Restoring controls to their original positions would restore the original colours, so they would understand they had caused the change, but the reason for the change would be inexplicable. This would violate one of the primary dictates of user interface design, predictability.

If the appearance of the characterised interface is typical of that expected when the interface is used, the ratios between the areas of the different interface elements are unlikely to change significantly. Most importantly, using fixed (but typical) values allows the hues of colour schemes to be changed (e.g. by rotating the colour molecule) without any apparently random changes to the colour scheme.

The abstract-to-perceptual mapping used in the prototype implementation

The method used to handle out-of-gamut colours resulting from the abstract-to-real transform in the Colour Harmoniser prototype was the simplest option, and the one that has the most predictable user-interface behaviour: altering only the out-of-gamut colours.

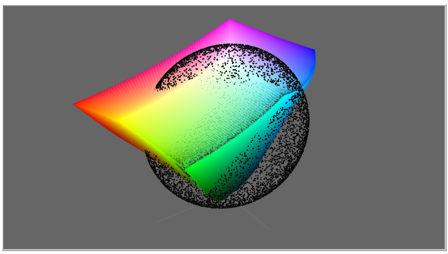


Figure 4.15: Colours inside the sRGB gamut for an LAB Scaling factor of 56. Only colours inside the dotted sphere can be present in the real colour scheme. This still excludes some high saturation colours, many of these can be included if the developer uses the personalisation controls (to be discussed later) to increase the colour intensity (saturation) of the recoloured interface. This will increase the abstract-to-LAB scaling factor which will increase the overall saturation of the scheme, but may force more colours out of gamut, which will require approximations to be made.

An illustration of proportion of colours available for use within colour schemes for a scaling factor of 56 (and those unavailable) is shown in figure 4.15. This scaling factor is sufficient to give reasonably strong colours, as can be seen by the colour schemes earlier in this section, all of which were created with an LAB scaling factor of 56. Strongly saturated colours for many hues are not obtainable with this scaling ratio, although in practice the lack of highly saturated colours does not appear problematic. The mapping from CIELAB space to sRGB used the *Little CMS* colour management system[23]. The behaviour of its transforms is such that, if the LAB scaling factor increases and pushes a scaled colour atom position further from the sRGB gamut boundary in CIELAB space, the resultant sRGB colour will also continue to increase in saturation, but with increasing hue error. The inaccuracies when the mapping reaches the gamut boundary (from theoretically correct colours to incorrect ones) is not obvious. Visually, the colours continues to change smoothly. Visual inconsistencies only become apparent when the error is extreme.

It is desirable to be able to adjust the saturation of schemes iteratively. To allow such changes, a method for the developer to control the saturation of colours in the final scheme is covered in the next section.

[23] `http://www.littlecms.com`, accessed May 21, 2010.

4.8 Colour scheme personalisation

When one is considering how the colours within a scheme might be adjusted, it is helpful to look at existing systems used for performing related tasks. Two widely used examples are colour television sets and more recently, photo-editing programs. The most basic and widely used colour adjustments are those for adjusting the colour saturation, correcting for apparent hue errors, and adjusting the contrast. All of these relate to the appearance of a visual image and how a user prefers to view it. While there may be, in an absolute sense, a "correct" colour value, controls are almost always provided to allow for personal preferences. It would seem reasonable for any automatic colour scheme generation system to provide the similar controls, so hue and saturation adjustment controls are provided by the Colour Harmoniser prototype. However, allowing the contrast to be reduced would reduce the light-dark difference and therefore the readability of text, so, unlike a photo-editing system, the Colour Harmoniser prototype has no contrast control: the limit values of the abstract colour space y axis (-1 and $+1$) are always mapped to black and white respectively (CIELAB $L*$ of 0 and 100).

4.8.1 Controlling the dominant hues in the scheme

It is possible to provide a hue (colour adjustment) control by incorporating a rotation about the y (lightness) axis when transforming the colour atom's abstract colour space coordinates into perceptually uniform space coordinates, where the rotation angle is controlled by the user's hue adjustment control. As the projected atom positions rotate around the lightness axis in the CIELAB space, so too will the sRGB colours resulting from the CIELAB to sRGB mapping. These updated sRGB colours are then used to recolour the interface elements associated with each colour atom.

The visual effect is that the hues of the interface colour scheme track the user's movement of the colour adjustment control. As the user adjusts the hue control, the colour scheme of the whole interface changes holistically. As illustrated in figure 4.16, a colour scheme can be changed from one based on red to one based on blue or purple or green, without altering the visual relationships between the interface elements. Colours that were desaturated remain so. Those that had high lightness values still have, but in (for example) green rather than red. Hue rotation is the first form of personalisation.

4.8.2 Controlling the colour saturation

Like the preferred hues, the appropriate saturation is a subjective decision made by the developer to include their preferences, and possibly those of the intended users (sec. 2.1.4). A control that allows the user to adjust the saturation of the final colour scheme without affecting either the hue or the light-dark contrast is desirable. With the saturation control at its minimum, the colour scheme will consist solely of black, white and shades of grey. When it is at its maximum, the colour atoms at the end of the wireframe will be mapped to highly saturated colours.

The saturation control is implemented by allowing the user to control the value of the LAB scaling factor that is used to map the chromaticity axes (but not the lightness)

Figure 4.16: From top left, the different colour schemes resulting from rotating a molecule in 30° steps about the lightness axis as it is projected in the CIELAB space.

from the abstract colour space to the CIELAB space. Figure 4.17 shows the effect of varying the scaling factor. The colours range from greys with a hint of colour to strongly saturated. If desired, all colour can be removed from the resultant colour scheme by setting the LAB scaling factor to zero.

This works well, but introduces the possibility of anomalous behaviour. A known fixed value of the LAB scaling factor was used during the optimisation to ensure a sufficiently large ΔE separation between the colours of items required to be distinguishable. Allowing the user to reduce the LAB scaling factor below the value used during the optimisation can reduce the actual colour differences between some interface elements. This can result in items being no longer sufficiently separated to be easily distinguishable. Nevertheless, if the user chooses to reduce the saturation and effectively waive one of their earlier distinguishability constraints, the Colour Harmoniser prototype will not prevent them doing so.

4.8.3 Allowing for the “natural” placement of colours

The third and final personalisation control is one that allows a complete scheme to have its lightness values inverted – light colours become dark and dark become light – without altering hue or saturation. This is to allow for the visual asymmetry in the appeal of certain colours. Some colours seem more suited to appearing in colour schemes when dark, and others when light. Yellow, for example, is most appealing when seen with high lightness values. The dark version of yellow is a khaki or brownish colour, and is much less common. The colours that are most naturally light or dark are fairly apparent from the shape of the Munsell or, (as shown in fig. 4.13 (p150)), the CIELAB colour space.

An optimised colour scheme rotated so that colour atoms with a yellow hue are darker than mid-grey (e.g. khaki) is unlikely to be appealing. However, inverting the lightness values of the colours, so that all the light colours become dark and all the dark colours (including yellow) become light, can result in a very acceptable scheme. Having the ability to flip the lightness values allows this possibility to be checked with no special effort on the part of the user and it accesses a set of colour schemes that have the same optimisation values as the parent (uninverted) schemes, but are very different visually and not available via the colour adjustment control[24]. The same inverted colour scheme could have arisen from the optimisation, but the reflected scheme might not be included in the set of ten presented to the user. This extra control ensures the reflected schemes are easily available at virtually no additional user interface complexity; only a single switch is required. The effect can be seen in the schemes in figure 4.18. The inversion can transform a usable, but visually unappealing, scheme into a colour scheme that combines usability and aesthetic appeal.

[24] The colour adjustment control rotates the colour atom positions around the vertical axis and so it cannot alter the lightness of the related interface elements. Whichever end of the tilted wireframe is lower than the $L* = 50$ (and the associated colour atoms) is not changed as the hue is changed. The interface elements linked to the colour atoms on the lower part of the wireframe will be always be dark. The ability to move colour atom positions vertically in the CIELAB space is the extra facility added by the “flip-light-dark” control.

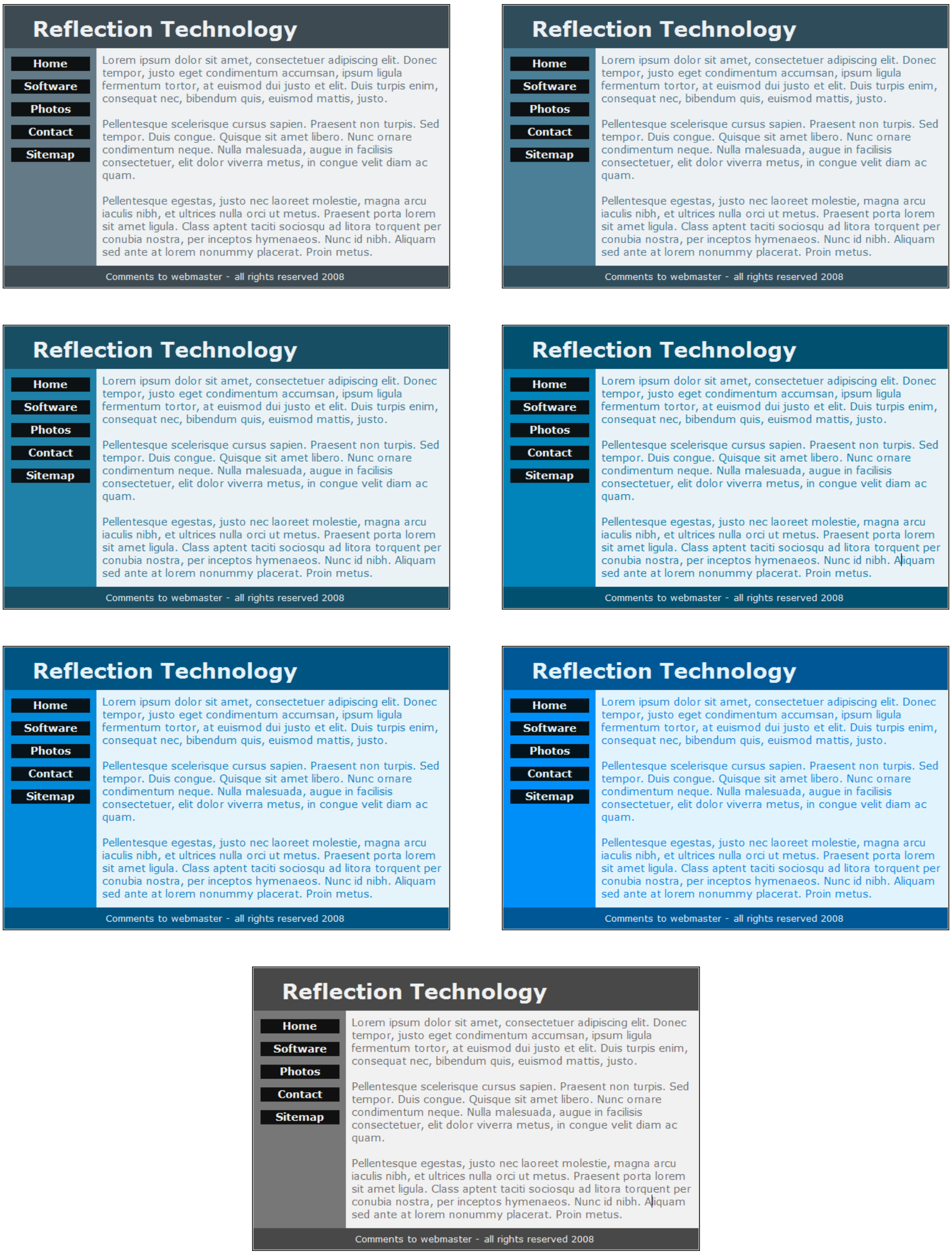


Figure 4.17: Varying the LAB scaling factor as the abstract colour scheme is projected in CIELAB space can act as a saturation control, enabling smooth control of the overall intensity of the resultant scheme. The six upper images show a monochromatic colour scheme of varying saturation, with LAB Scaling factors of 15, 30, 45, 60, 90, 150. The lowest scheme has a LAB Scaling factor of 0. This removes all colour, without impairing usability.

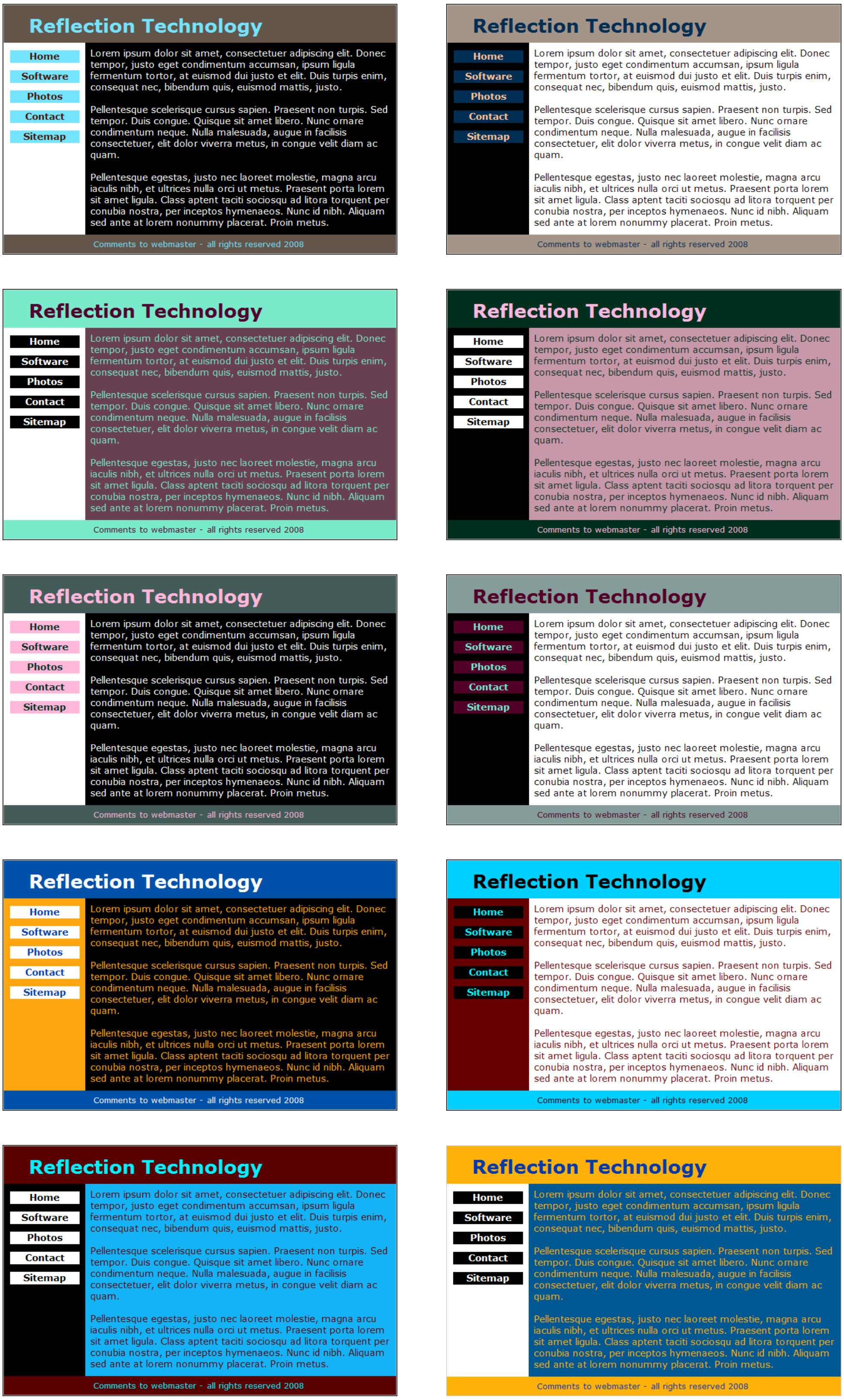


Figure 4.18: Each row shows the same colour scheme, but one of the pair has the lightness inverted, which changes the colours without affecting the colour relationships. The schemes are based on a split-complementary wireframe including the black-white axis. The top six schemes have a LAB scaling factor of 56, the lower four ≈ 145.

The nature of the personalisation (hue/saturation/flip light-dark) controls ensure that both during and after adjustment, the harmonious nature of the colour scheme is maintained and the readability and user-defined aesthetic constraints[25] are preserved.

4.8.4 Implementing the personalisation controls

These three means of personalising a colour scheme – altering the hue, controlling the colour saturation, and flipping the light-dark colours – allow the developer to explore a large number of variations of the schemes created by the optimiser.

The user controls used to provide these personalisation capabilities are intended to be sufficiently simple to require virtually no explanation and still yield a very large number of colour scheme variations from the same abstract colour molecule.

The transform used to implement the personalisations is:

$$p_i = kFRa_i \text{ — applied to each of the atoms } a_i \text{ in a colour molecule} \tag{4.23}$$

where:

p_i – perceptually uniform colour space coordinates

a_i – $(x_i, y_i, z_i)^\mathsf{T}$ — abstract space coordinate of atom$_i$

k – "Colour Strength" (saturation) control value — an integer

R – $\begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$ — 3D hue rotation

θ – "Change Colour" (hue) control value (in degrees)

F – $\begin{bmatrix} 1 & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$ — Lightness inversion

f – "Flip Lightness" $\begin{cases} -1 & \text{if control set to true} \\ +1 & \text{if control set to false} \end{cases}$

Points p_i in the perceptually uniform colour space (CIELAB in the Colour Harmoniser prototype) can then be transformed to the sRGB colour space. Despite the mathematical complexity of the transforms, from a user's perspective, the controls for manipulating the colour scheme are easy to understand: there are two familiar analog controls for adjusting colour and saturation, and one switch that can be used to see if the colours look better "the other way around".

[25] except where the user chooses to override their earlier choice.

4.8.5 Direct manipulation colour scheme exploration

When any colour molecule is selected, the transforms:

$$\text{colourAtom}_i \rightarrow \text{ CIELAB } \rightarrow \text{ sRGB}$$

are recalculated for all atoms in the molecule, and the interface display is updated with the new element colours. The hue, saturation and flip lightness personalisation controls are "live" – changing any of these controls also causes an immediate update of the target interface colour scheme. This provides immediate feedback on the appearance of the interface with the updated colour scheme, without the mode change and visual discontinuity encountered when using conventional modal one-at-a-time colour selectors that force users to choose colours out of context based on a small swatch.

Neither of these interruptions occur when adjusting a scheme using the colour personalisation controls. As well as the examples of the web page already illustrated, figures 4.19 and 4.20 illustrate the types of schemes that the Colour Harmoniser is capable of producing for application software and presentation slides.

4.8.6 Colour Harmoniser implementation – summary

The prototype enables a developer to choose an interface to be coloured, to specify constraints on the colouring and then to create a set of optimised colour schemes. These schemes can be evaluated and modified in real time, allowing the developer to examine a large number of colour schemes variations based on prototype schemes created by the optimiser.

An overview of the process and flow of data is illustrated in figure 4.21 and is explained in detail below:

1. *the interface characterisation phase (1, 2 in the diagram):* this takes data from two sources, the user interface to be coloured and the developer. From the user interface, pertinent details of the interface elements are extracted (e.g. interface element names and areas). The names of the interface elements are presented to the developer who can then group the elements, set the distinguishability constraints, and choose a particular colour scheme wireframe. This phase produces a list of colourable items (the colour atoms), the wireframe desired by the user, a list of text items and their backgrounds (where the text must be legible against the background), and a list of pairs of items that must be mutually distinguishable.

2. *construction of the colour molecules (3):* multiple colour molecules are constructed by choosing colour atoms positions corresponding to different ordering of atoms on the wireframe. These molecules are likely to have widely varying fitness.

3. *the optimisation phase (4, 5, 6)* this uses (i) the readability constraints list, (ii) the distinguishability constraints list, (iii) the selected wireframe, and (iv) colour balance criteria (not shown), to reposition the atoms within each colour molecule to improve its fitness score.

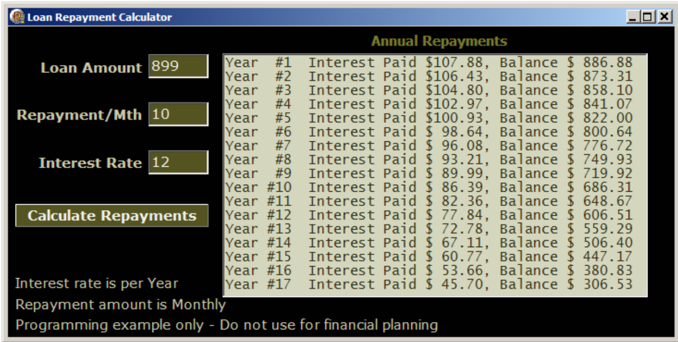
4. *display a colour scheme (7, 8):* the user selects one of the optimised colour molecules. Its colour atom coordinates are transformed from abstract colour space to sRGB and the resulting colours are used to recolour the target interface so the developer can evaluate the updated colouring.

5. *colour scheme adjustment (9):* the developer can change the colour scheme in a holistic manner by adjusting controls to alter the hues, the saturation, and whether the interface elements are shown with light or dark colours. Adjusting any of these controls will alter the parameters used in the abstract-to-real colour space transform, but will not alter the underlying relationships between the coloured atoms. All the optimised constraints are preserved, avoiding the need to reoptimise. Therefore, when the user adjusts any of the colour adjustment controls, the user interface colour scheme can be immediately updated, allowing a rapid and continuous exploration of many different colourings of the interface.

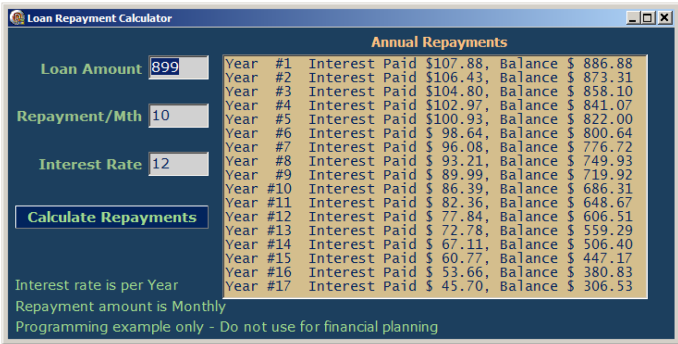
Several colour molecules are produced as a result of the optimisation phase, each of which will use colours from the currently selected wireframe in a different way. The user can select one of the optimised schemes and, by adjusting the holistic personalisation controls, explore and fine-tune different colourings. If, however, it is apparent that the distinguishability constraints or groupings are inappropriate, these can be altered and a new set of colour molecules generated, which can then be fine-tuned. The process therefore similar to working with a colour design consultant, but the process is interactive and the iteration time is shorter.

The aesthetic appeal and usability of any colour scheme generated by this process are primarily ensured by using the fitness function to evaluate and direct the choice of colour schemes. The validity of the four-factor based fitness function has not yet to be established.

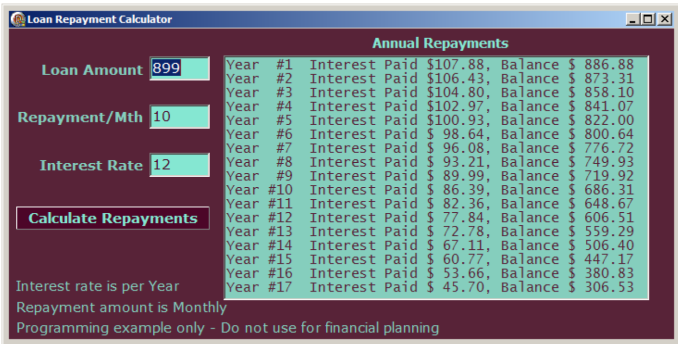
With a trial implementation having been implemented in software and available for use, it is possible to create optimised colour schemes and subject these to user evaluation. This will enable the experimental validation of the fitness function – a critical component of the architecture. This evaluation is the subject of the next chapter.



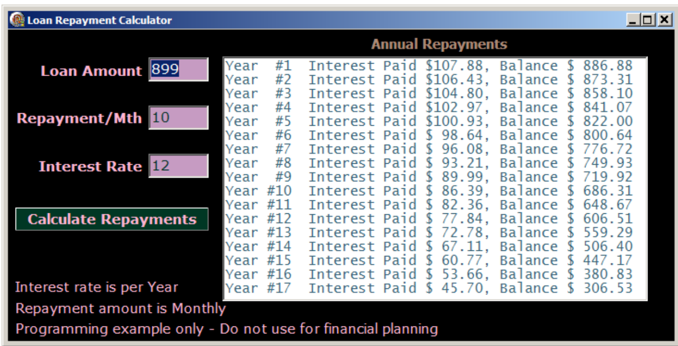
A monochromatic colour scheme



A split-complementary colour scheme

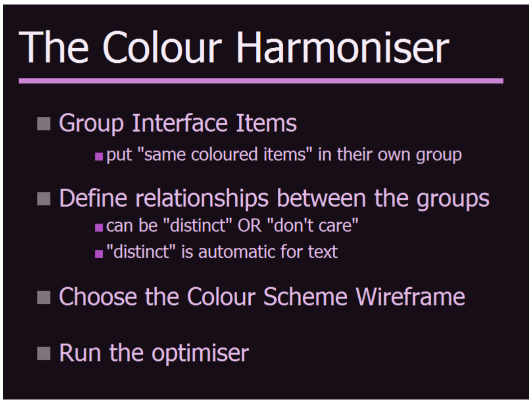


Too much colour, in a complementary scheme.

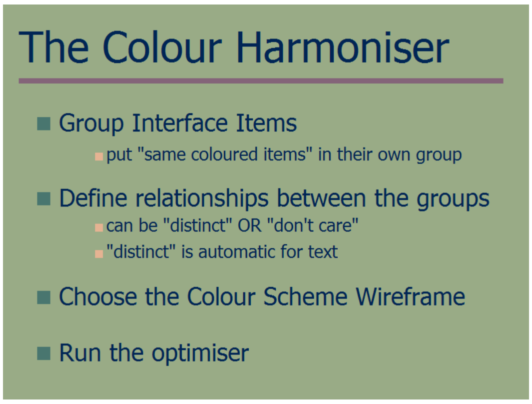


An elliptical + black/white colour scheme.

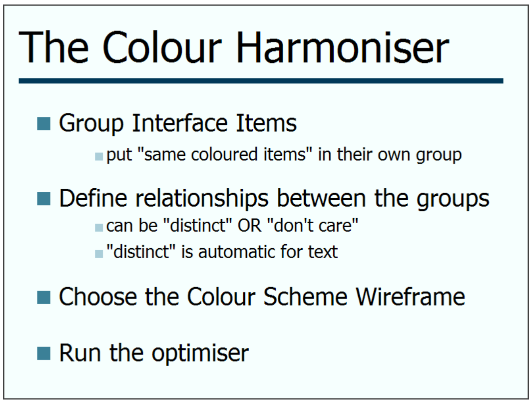
Figure 4.19: Desktop application colour schemes created by the Harmoniser prototype. Extending the wireframe model to include the black-white axis (e.g. the bottom scheme) allows the use of less colour, and gives more familiar schemes.



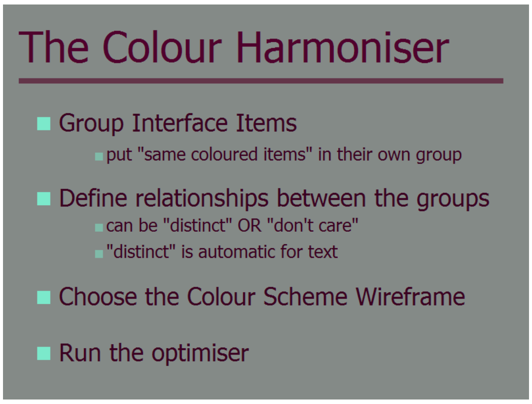
A monochromatic colour scheme.



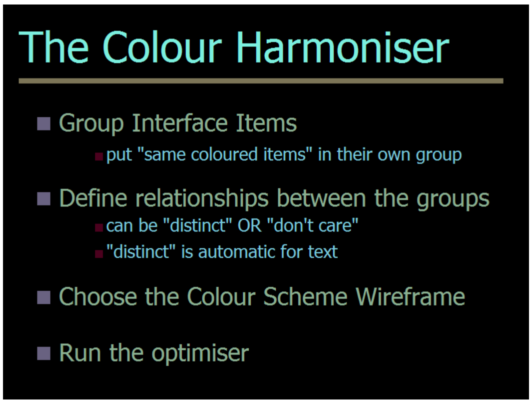
An elliptical colour scheme.



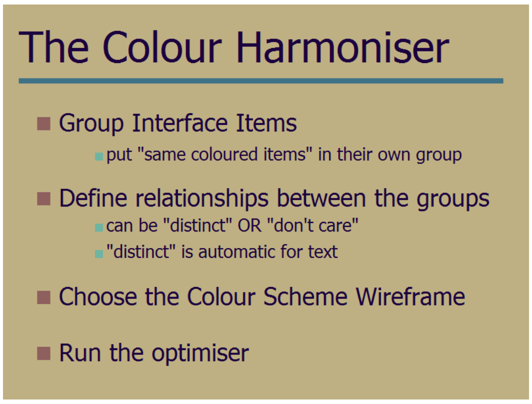
An analogous colour scheme.



A complementary colour scheme.



An elliptical + black/white colour scheme.



An elliptical colour scheme.

Figure 4.20: Colour Harmoniser-adjusted colour schemes for presentation slides. Almost all the area (∼90%) of this interface is background), so schemes that balance around mid-grey (e.g. complementary/split-complementary/elliptical) can only balance by placing the background very close to the origin (and it is therefore greyish) to minimise its effect. If the "use black-white" option is enabled (e.g. the bottom-left scheme), the background can be placed on the achromatic axis, and the smaller objects balanced against one another. It would seem that using area balance to determine the lightness and saturation of the elements is better suited to interfaces whose elements have a more even distribution of areas (or can be made to have this by grouping).

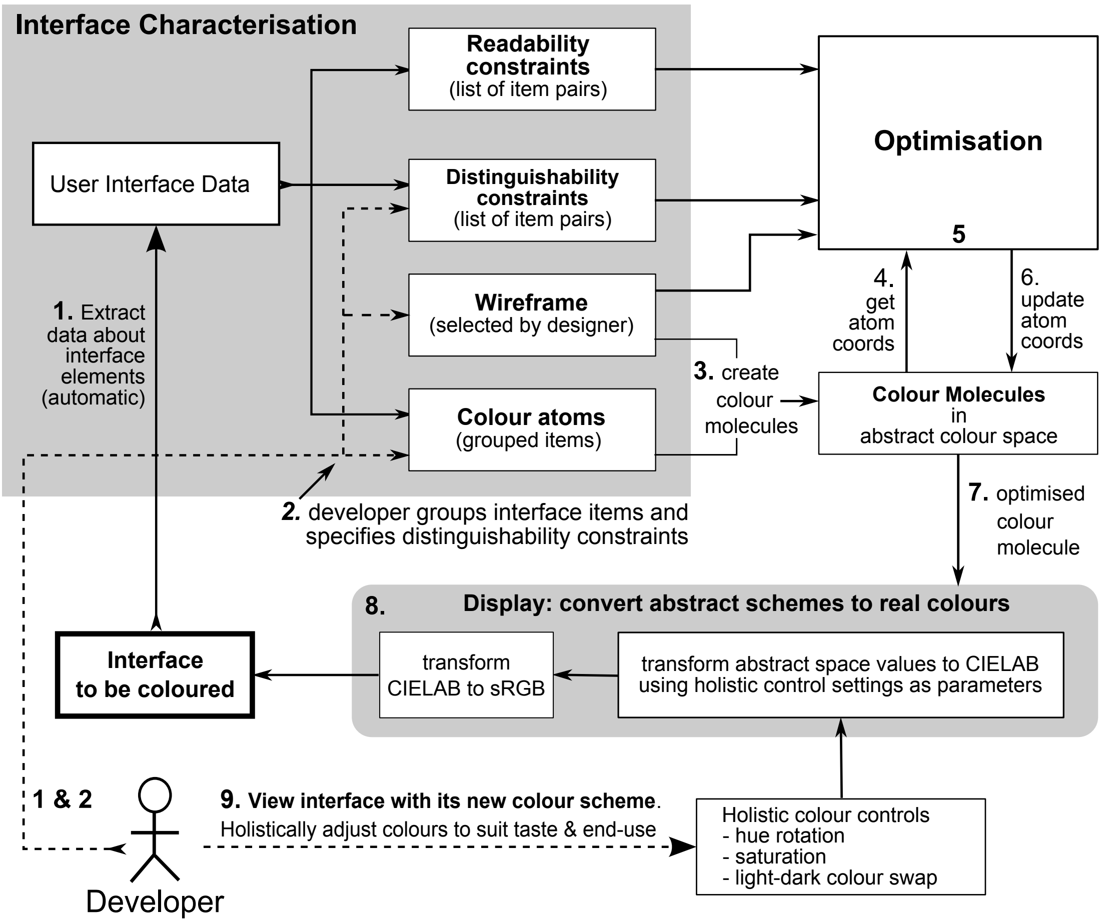


Figure 4.21: The data flow within the prototype Colour Harmoniser, starting with extracting data from user interface to be coloured and the user's constraints, through to the updating of the same interface with an optimised colour scheme. The solid lines indicate data flows within the software, the dotted lines are user inputs.

Chapter 5

The experimental validation of the fitness function

Although the main experimental evaluation of the Colour Harmoniser prototype is in the next chapter, some preliminary experiments were carried out to assess the suitability of the fitness function and its parameters. These experiments and their results are the subject of this chapter.

There are factors that are common to all the experiments, so these will be covered here, rather than at the beginning of Chapter 6. Later, when each experiment is described, only the points of difference will need to be included.

5.1 An overview of the experiments and methods

To clarify which experiments were performed and why, prior to encountering the detail of each one, the experiments are summarised in Tables 5.1 and 5.2.

All the experimental trials were computer-based and started with a page that introduced the experiment, what it was for, who was running it and why, and what the participant would have to do. This was followed by a demographic survey that included questions relating to the participant's previous colour and design experience and colour-related self-perception. The body of the experiment followed, generally followed by a concluding questionnaire and a "thank you" page.

Ethical approval: The experimental work undertaken for this thesis was approved by the Massey University Human Ethics Committee, approval #04-184.

5.1.1 The evaluation interface used in the colouring trials

This thesis aims to determine whether an extended set of colour harmony guidelines can be used to create good quality interface colour schemes. To test this experimentally requires the participants in the experimental trials to either colour or evaluate an interface.

Overview of Experiment Set 1: Validating the Fitness Function

Aim	Overview
Pilot Fitness Function Validation Experiment: To determine whether or not the Colour Harmoniser's fitness function can operate as an adequate alternative to human assessments of colour schemes.	*Method:* Participants ranked colour schemes; these rankings were compared with a ranking derived from the fitness scores. The schemes within each set were related, as they were saved during the optimisation of a single colour scheme. *Subjects:* 12 females, 11 teenagers/1 adult. *Location:* University Lab
Main Fitness Function Validation Experiment: To determine whether or not the Colour Harmoniser's fitness function can operate as an adequate alternative to human assessments of colour schemes. This experiment is largely a repeat of the previous experiment using a more representative population. The data collected was also used to evaluate appropriate weights for the fitness function terms.	*Method:* Participants ranked colour schemes; these rankings were compared with a ranking derived from the fitness scores. The schemes each set were chosen randomly (one of rank_1, one of rank_2 etc.). Therefore, the schemes within each set are not related, unlike the previous experiment. *Subjects:* 19, 10 female, 8 male, one unspecified, aged 15-65+. *Location:* University Lab, UCOL Design School, Museum and Art Gallery, homes and businesses.
Establishing the Independence of the Fitness Function Terms: To determine whether, for high scoring colour schemes, the four fitness function terms are independent.	*Method:* A Monte Carlo method was used to create a set of high scoring colour schemes. The correlations between each of the four terms was then calculated to determine any dependency.
Estimating the fitness function term weights from experimental data: To use the user evaluations of colour schemes to estimate an improved set of fitness function term weights.	*Method:* Regression was used to find the coefficients for each of the terms that would give a fitness score that best approximates the colour scheme assessments of the participants in the Main Fitness Function Validation experiment.

Table 5.1: Three sets of experiments were used to evaluate the viability of the Colour Harmoniser approach to colour scheme design. The first set (shown in this table) are oriented towards validating and adjusting the parameters of the fitness function. The next two (Table 5.2) assess the usability of the Harmoniser method and the quality of its schemes.

Overview of Experiment 2:	**Compare the Methods of Scheme Creation**
Aim	Overview
To compare the Harmoniser-based methods of colour scheme creation with conventional methods by comparing the usability of the methods, the quality of resulting schemes, and the time taken to produce them. A side effect was to collect a set of humanly-adjusted Harmoniser-based schemes.	*Method:* The participants created two colour schemes, one using a conventional method and the other by selecting and adjusting a Colour Harmoniser-created colour scheme. On completion, the quality and usability evaluations were assessed, both for each method and comparatively, using Likert scales. *Subjects:* 73 (39 female and 34 male). *Location:* University Labs and homes.

Overview of Experiment 3:	**Compare the Quality of the Results**
Aim	Overview
To evaluate the quality of colour schemes created by the Colour Harmoniser and several other methods using several relevant quality criteria, to allow a comparison between Harmoniser-based and human-created schemes.	*Method:* The participants were shown colour schemes individually and asked to score each scheme on four criteria using Likert scales. Five different categories of colour schemes were represented. The schemes include those by: artists, non-artists, the Colour Harmoniser before adjustment (raw), Colour Harmoniser-adjusted, and schemes using random colourings. The images were selected randomly, but with mechanisms in place to ensure an even distribution across all five methods. *Subjects:* 127, with a balanced gender ratio (50%:50%), and ages ranging from 5–14 to 65+. *Location:* University Lab, UCOL Design School, Museum and Art Gallery, homes and businesses.

Table 5.2: An overview of the second and third experiments used to evaluate the viability of the Colour Harmoniser approach to colour scheme design. The second experiment assesses the usability of the Harmoniser method of colour scheme selection and adjustment in comparison to more conventional methods of colour scheme design, while the final experiment compares the quality of colour schemes produced by the Colour Harmoniser (both before and after adjustment), with schemes created by people.

This section discusses the properties that influenced the selection of the interface whose colour scheme was evaluated in the experimental trials. It is not possible to use an "ideal" or an "abstract" interface in the user trials: realistic interfaces must be chosen, preferably ones that are familiar and simple enough to be effectively invisible, so the evaluators are not distracted by the interface itself. It is clearly not possible to test all interfaces, so one or more illustrative interfaces must be selected for use in the experimental trials.

Everyday computer users may wish to colour desktop applications, websites, documents or presentations. The creation of desktop applications is rare, and the colouring of documents is not particularly useful, as almost all non-presentation-oriented documents are black and white. Web sites and presentations are similar, but web sites have a more complex structure.

Web sites have a fairly uniform architecture, due to the need to identify the current site for the user, and to provide navigational controls and content. An example of such an interface is the prototypical web page used to illustrate the colour schemes in the last chapter (e.g. fig. 4.16). This layout is sufficiently familiar, even to non-specialist computer users, to need no explanation.

Desktop application interfaces with a menu bar at the top and the status bar at the bottom are common, but the content of the area between the menu and status bar varies, usually containing a single large body area, or a body area with one or more navigation panels and toolbar buttons. There is little standardisation.

A typical presentation (e.g. a *PowerPoint*™) slide has a title at the top, the corporate affiliation and page number at the bottom, body text and the slide background. Small decorations, such as coloured graphics for bullets, are widely used. Overall, the structures of both desktop applications interfaces and presentation slide are quite similar to the web page already discussed, except that with presentation slides the fonts are bigger and there is no navigation panel.

Interface elements	No. of colours for individual colouring	No. of colours with grouped elements
page body: text and background	2	2
header: text and background	2	2
footer: text and background	2	(same as header)
navigation panel background	1	1
5 × button text colour	5	1
5 × button background	5	1
Total number of colourable elements	17	7

Table 5.3: A list of the colourable elements on the prototypical web page (fig. 4.1, p115). The middle column shows the total number of colours if each element can be coloured differently. The right-most column shows the effect of grouping related elements and using the same colour for the elements of a group. The number of colourable elements drops significantly, and the use of common colours for multiple interface elements also increases the visual coherency of the scheme.

The prototypical website reflects the layout and elements to be found in a large number of graphical user interfaces. The number of colourable interface elements (seventeen – see Table 5.3) is sufficiently large to make creating a pleasing colour scheme challenging for the non-specialist. However, using different colours for each of the buttons and its text would look odd. More usually, the text on all the buttons would be the same colour, and the buttons themselves another single colour. This reduces the number of colours to nine, and using a common colour for the header and footer background, and another for the header and footer text, reduces the total number of colourable elements to seven, as shown in the rightmost column in Table 5.3.

As has been shown, the Colour Harmoniser prototype (and therefore the method) can be used to produce harmonious and usable colour schemes for interfaces other than the prototypical website, but given the generic nature of a website interface, its familiarity, the sufficiency of elements to colour, and the need to keep several items of text distinct from their backgrounds, the prototypical web site interface has been selected as the interface to be coloured and evaluated by the participants in the experimental trials.

This research was centred on developing a colour harmonisation technique that can control the colours of all of the components of an interface; it was felt that including graphics and images – whose colours are typically fixed – in the test interface would make it more difficult to analyse the experimental results. Therefore, pre-coloured graphics and images were not included in the test interface.

The elements used in the test interface are each coloured with a single solid colour. This is a widely-used method of colouring user interfaces and is not unduly restrictive. If the Harmoniser-based method of colour scheme creation is successful, follow-up work could focus on the integration of features such as pictures, graphics, textures and transparency. Some ideas relating to the inclusion of pictures and graphics, and how their presence may be integrated into the Colour Harmoniser method are discussed in section 7.5.

Colouring an interface with only seven colourable elements is not a trivial problem. The unconstrained selection of colours using the ubiquitous 24 bit colour selector for seven colourable elements yields $16,777,216^7$ possible colourings. The colouring of the prototypical web page is a realistic task. It includes the elements typically found on web pages (see Wu and Chuang (2000) and Table 5.4), and has sufficient colouring complexity to justify its use in the experimental trials.

The evaluation of human colour preference is subjective, and there are a very large number of possible colour schemes, therefore it is important to use as large a number of samples as possible. The use of a single interface removes the type of interface as a possible confounding factor, and for a given number of samples, enables the most robust statistical analyses. As has been demonstrated (figures 4.19, p164 and 4.20, p165), the Colour Harmoniser can produce colour schemes for application software and presentations, and if the results of these experimental trials described here are favourable, confirmation using other styles of interface would then be warranted.

Design elements	Frequency	Percentage
buttons*	50	100%
counter	42	84%
title*	40	80%
content*	39	78%
picture	27	54%
background graphics	26	52%
background colour*	24	48%
flashing light	21	42%
date	16	32%
search engine	9	18%
* = elements in the test interface.		

Table 5.4: A breakdown of the interface elements found in a survey of fifty personal web sites by Wu (2000). The elements in the prototypical web page used in the experiments are indicated with an asterisk: all the elements with essential semantic content are included.

5.1.2 The inclusion of CVD participants in the experimental trials

Participants in the experimental trials were not screened for colour vision deficiencies, but are asked whether (as far as they know), they had normal or impaired[1] colour vision. Those who stated that their vision was impaired were not excluded from participation, but the responses to the question were recorded in the data from the trial.

The lack of a comprehensive test for normal colour vision might be thought unusual, but, as will be explained shortly, the inclusion of such screening would offer minimal benefit. The requirement to screen each participant would have added another stage to the experiments. Most already require the participants to understand multiple screens of information. Increasing the number of screens would have made the experimental trials more time-consuming and so, to maximise participation and completion by the participants (all of whom were unpaid volunteers), CVD screening was not included. The presence of a small number of participants with colour vision deficiencies is not expected to significantly affect the experimental results. This lack of effect was expected as

1. there is a large difference in lightness between any text and its background in colour schemes created using the Colour Harmoniser method, and so its colour schemes are readable by those with colour vision deficiencies. Hence, their evaluation scores should not differ from the norm because of indistinguishable text. The experimental results (which can only be indicative as there were few CVD participants) did not shown any CVD effect in the assessments of Harmoniser-based schemes (Tables 6.5, p228 and 6.9, p230).

2. the colour schemes created by those who indicated they had impaired colour vision, or chose to skip that question in the survey, were not used as samples of

[1] The options are *normal* or *impaired (i.e. you are colour-blind.)* This is sufficient for a first-order analysis without requiring details of the degree of impairment, which would require specific user screening.

humanly-created colour schemes in the *Compare the Results* experiment. This removes the possibility that any unusual schemes created by those with atypical colour vision would appear as outliers in the image pool.

3. the *Compare the Methods* experiment included slightly more females than males (39:34), and in the *Compare the Results* experiment, the genders of the participants were evenly split. As colour vision deficiencies are rare among females (see sec. 2.3.2), the number of CVD participants is unlikely to exceed 4%, so any disruption of the statistics is likely to be minor.

While no significant differences are expected from colour vision deficient users when viewing Colour Harmoniser-created colour schemes, the collection of data from CVD participants allows their evaluations be compared with the non-CVD participants. Such results are reported in the next chapter during the analysis of the *Compare the Results* experiment.

5.1.3 The participants in the experimental trials

Over three hundred unpaid volunteer participants took part in the trials, in a variety of locations: passers-by and visitors to a museum and art gallery, a design school foyer, University labs, a Linux user-group meeting, businesses, and homes.

Participants were volunteers and therefore self-selected, and while there was no explicit matching of demographic categories (e.g. age, gender, colour design background), the variety of sites chosen for data collection helped ensure a wide sampling within the categories. The demographic summary of the "Compare the Results" experiment (in Table 6.1) illustrate that a broad sampling was achieved, with two exceptions: ethnicity and colour vision deficiencies. Most of the participants were European[2] and the numbers in the other groups were very small, so no analysis using ethnicity is included. Only 3% of evaluations were from participants who indicated a colour vision deficiency; however this is as expected[3]. Therefore, although results testing for CVD effects are presented, they should be taken as indicative only.

5.1.4 The experimental conditions

The experimental trials were undertaken in a variety of locations with environmental conditions typical of those encountered by home users and non-professional developers. Strict control of lighting was not possible; but it was always indirect, and the most common conditions were either daylight or a mixture of daylight and artificial illumination. However, while the lighting could not be controlled, the type of lighting in use was noted as part of the experimental results.

As the intention in the experimental trials is to assess the harmony of complete schemes, not colour matching in an absolute sense, strict environmental controls are

[2] African 1%, Asian 7%, European 84%, Latin American 3%, Maori 3%, Middle Eastern 3%.

[3] For a sample with the genders evenly represented, about 4% of the participants would be expected to have colour vision deficiencies (Fairchild, 1998).

not required and, as noted by Laugwitz (2001), the results of colour harmony experiments under more controlled experimental conditions could be expected to yield similar results. Conducting the trials in uncontrolled, but typical, environments helped to ensure that the experimental results would reflect the experiences of typical users creating or viewing Harmoniser-based schemes.

The experimental machines used in the experimental trial all used modern LCD panels, and were running Microsoft *Windows XP*™. In the first experiment (*Pilot Fitness Function Validation Experiment*), the default system sRGB colour profile was used. For the later experiments, the system colour profile was set to an ICC sRGB colour profile created by a GretagMacbeth *EyeOne-Pro*™ spectrophotometer.

5.2 The experiments used to evaluate the fitness function

The final stage in colour scheme creation using the Colour Harmoniser method is the fine-tuning of candidate schemes produced by the optimiser. It is important that the schemes presented for fine-tuning are "good" colour schemes, as the fine-tuning can only alter the overall colours used, not the underlying structure of the colour scheme. Therefore, during the creation of the candidate schemes, the Colour Harmoniser must be able to choose good schemes. It must be able to assess the merit – the fitness – of one colour scheme over another, as this will enable the selection of suitable schemes from the myriad of possibilities. It is important that the ratings from the function evaluating colour schemes fitness should agree with those ratings produced by humans. This will enable the fitness function to act as a proxy for a human evaluator during the selection of colour schemes.

The fitness function is based on artistic criteria for colour scheme design, augmented for factors relevant in interface design. It is not possible to assert (or require) that the fitness function assesses colour schemes "in the same way humans do", as how humans assess overall colour schemes is unknown. The best we can hope for is a strong correlation.

It would be sufficient if the ordering of schemes scored highly by fitness function corresponded to the ordering of schemes scored highly by the human evaluators, as this would allow good schemes to be chosen based on the fitness function score. However, it is preferable that this ordering extend to schemes with low fitness function scores, as this would allow gradual improvements in a scheme to be reflected in an incrementally increasing fitness score, and would allow the function to be used as part of an optimisation process. This is important as the initial placements of colour atoms are very unlikely to have high fitness scores, and these low scoring arrangements are the starting point for the optimisation.

To establish a relationship, the colour schemes could be scored or ranked. Scoring requires the evaluation of individual colour schemes against an external standard: the user is comparing the presented colour scheme with their own personal standard of a "good colour scheme". A trial based on ranks is self-referential, as each scheme is being compared to other schemes in the same set. Therefore, a high rank does not indicate (unlike when scoring a scheme on a “terrible” to “excellent’ scale) that the evaluator thinks the scheme is a good one. A highly ranked scheme may, in the evaluator’s estimation, be poor, but not as poor as the lesser ranked schemes. Ranking does however allow a large number of schemes to be evaluated, as ranking one set orders multiple schemes.

5.2.1 Assessing the relationship between the fitness function score and human colour scheme assessment

As the fitness function score is the basis for selection of good colour schemes within the Colour Harmoniser method, for the method to work, the evaluation must be sufficient. That is, the schemes that it scores highly must also be assessed highly by human evaluators. There is likely to be significant variability in the human evaluations of any given colour scheme, but nevertheless, averaged over multiple evaluators and multiple colour schemes, agreement between humans and the fitness function assessment is essential.

A moderate or strong correlation may exist between human evaluators and the fitness function, even if important components are missing from the fitness function. For example, a scheme that is scored highly by the fitness function may not be regarded as a good colour scheme by humans. This could lead to the situation where the fitness function and people have a high correlation, but the correlation is for part of the quality scale. For example, on a quality scale of one to ten (the best), the factors being considered may result in schemes that can only score a maximum of six out of ten, and some other criterion needs to be added to reach a score of ten. Therefore, to establish the sufficiency of the fitness function, it is necessary to both show a positive correlation between the colour scheme evaluations by the fitness function and by human evaluators, and to show that the final colour schemes produced using the Colour Harmoniser method are scored highly by human evaluators. Each part is addressed separately. The first (establishing a positive human/fitness function correlation) is evaluated in two experiments; the *Pilot Fitness Function Validation Experiment* (below) and the *Main Fitness Function Validation Experiment* (sec. 5.2.3, p182). The second, (determining if high scoring Colour Harmoniser schemes are good user interface colour schemes) is tested experimentally in the *Compare the Results* experiment, detailed in the next chapter.

5.2.2 A pilot fitness function validation experiment

Aim of the experiment

To determine whether or not the Colour Harmoniser’s fitness function can operate as an adequate alternative to human assessment of colour schemes.

Experimental design

Both scoring and ranking require colour schemes of widely varying quality in order to establish a range, or a basis for the rankings. This introduces a potential difficulty into the creation of the set of colour schemes for the trial. In the absence of any standard method of measuring colour scheme quality, how are these colour schemes selected or generated? This is no simple solution, but fortunately ranking does not require an unknowable standard. The schemes to be ranked can be created using the fitness function with the fitness scores to allocate the schemes to ranks.

It was decided to use a trial based on the ranking of colour schemes within a set as:

- ranking gives an automatic basis for comparison, rather than some personal and idealised standard.
- with ranking it is possible to get the subjects to assess more images[4] than would seem reasonable with individual image evaluations. In informal pre-trial experiments it was found that ranking 16 sets was not too taxing. This gives 96 assessed images per subject. The user interface for the ranking-based trial showed progress as "6 out of 16", whereas using individual evaluations this would have been "36 out of 96". Even though the number of remaining of assessments is identical, having 60 more images to score sounds much worse than having 10 sets. As the participants were unpaid volunteers, it was thought more likely that participants would complete all the image sets in a trial if the ranking method was used.
- on a 17" display (the smallest used in the trials), the colour detail from the individual colour schemes can be seen without difficulty when six colour scheme images are displayed simultaneously.
- six images per image set is sufficiently few to allow significant visual differences between the members of the set.
- the chance of a single random ranking of six colour schemes matching that of the fitness function is less than 3%.

Experimental method

Participants are required to rank sets of six website colour schemes according to the their assessment of the use of colour within the scheme and its readability. Each image set has one image from each of the six different bands of fitness as illustrated in figure 5.1, so there is a significant visual difference between the images within a set. Due to the design of the experimental interface, the participants cannot assign tied ranks. The order of the images within each set was varied randomly.

Creating the ranked sets of images The fitness function score is continuous. To ensure the images in each set had visually distinct differences, it is desirable that the fitness scores for images in one band of fitness are significantly different from those in the adjacent bands. This was implemented by imposing artificial fixed boundaries in order to allocate each colour scheme to one of six categories, as shown in figure 5.1. The schemes within each image set are chosen from a single population and

[4] Both "colour scheme" and the shorter "image" are both used to refer to a colour scheme being ranked by human participants.

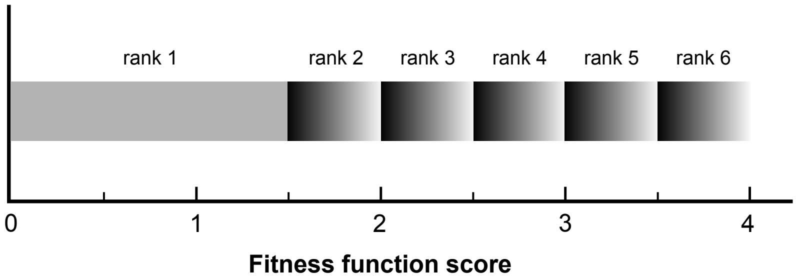


Figure 5.1: The colour schemes to be ranked by the participants were chosen from six different bands of fitness. The lowest scoring colour schemes in a population are used as the $rank_1$ sample. The $rank_6$ scheme was the best at the end of the optimisation, with a score of at least 3.5, with the remaining four being the first schemes during the optimisation to exceed the fixed fitness thresholds of 1.5, 2, 2.5 & 3. The darkness of the shading indicates the region more likely to have samples in each rank band.

include the scheme with the lowest fitness from the initial population, the scheme with highest fitness after the optimisation, and four schemes of ascending fitness during the optimisation. The steps in the fitness score result in significant differences in the perceived quality of the schemes in the different ranks. This helps mitigate the effect of the forced choice ranking (ties are not permitted).

The Munsell Method of Colour Balance assessment was used
The colour balance term in this trial (for the complementary and split-complementary colour schemes) was assessed using the Munsell law of colour harmony, balancing area × saturation × lightness, with all wireframes at a fixed tilt of 45° and an LAB scaling factor of 56.

This data set for this experiment consisted of randomly selected schemes created by the Colour Harmoniser. Four schemes were randomly selected from those generated using the complementary, split-complementary, elliptical and monochromatic wireframes. This resulted in sixteen widely varying colour schemes because of the differing wireframe shapes and randomly selected hue rotations.

Despite the forced randomisation, within each set, the colour schemes have some internal coherence as the schemes are sampled during the optimisation of a single population. In the early stages of optimisation, the images within the population are quite diverse and include the random schemes from the initially seeded population. Visually, these can be quite different and have no apparent relationship to more optimised schemes. As the optimisation proceeds, schemes with low fitness (which usually include the random schemes) disappear from the population. This reduces the variability in the population, and the optimisation changes become more local. The visual effect is that of the colour scheme stabilising, with the higher ranked schemes within a set being visibly similar. As all the schemes within a set have the same wireframe rotation (and therefore use the same hues), this helps eliminate the actual colours used in the scheme as a confounding factor. Such an effect might occur if the schemes used different colours: an evaluator might favour a scheme if it used their preferred colours rather than considering the overall colour scheme.

The experimental task The sixteen sets of colour schemes were shown to each participant. Each image set contained six images, one colour scheme from each fitness band as described above. The initial ordering of images within a set was varied randomly. The participants were instructed to click on the scheme that combined the best use of colour with good readability. This top-ranked image would then disappear, leaving only the lower (and as yet) unranked schemes. The user then clicked on the best remaining scheme, which would then vanish. This continued until all the schemes within the set had been ranked.

The software would then redisplay the same set ordered according to the user's indicated preference. If the user was happy with the revised ordering, they requested the next set, if not, they could reorder the set as described above. No time or iteration limit was place on users. Each participant ranked the same sixteen sets of six schemes.

Subjects and experimental conditions The participants in this initial trial were from a local school – eleven female students aged thirteen to seventeen and their teacher, also female. The experiment was performed in a lab at the university using identically specified Windows XP machines with Phillips 170B 17" LCD displays set to their sRGB setting. The lighting was a mix of daylight and overhead fluorescent lighting.

The participants were all female and primarily teenagers. This restricts any general conclusions that can be drawn, but the results were intended to give an indication of whether a more balanced trial was warranted.

Results

Each of the twelve participants ranked sixteen sets of colour schemes, each with six colour schemes per set. This resulting in 192 orderings. The orderings of colour schemes within each set from the participants was compared with the 6-5-4-3-2-1 ranking based on the fitness function score using the Spearman rank correlation coefficient[5]. The results were in general agreement, as can be seen from the distribution of the Spearman coefficients (ρ) shown in figure 5.2. To assess the statistical significance of the agreement, the correlations were assessed by image set, by participant and overall. Coming from human participants, these are strong correlations (Meyer et al., 2001).

Correlations between users and fitness function, by image set

To determine if the correlations differed between the image sets, the evaluations for each image set were assessed separately. The results are shown in table 5.5.

[5] Rank correlations could be compared using any one of several methods, including the Kendall-τ correlation coefficient and the Spearman rank correlation coefficient. The Spearman correlation coefficient was chosen as this is the most widely used method of comparing ranks, and therefore simplifies interpretations of the results and the comparison of the degree of agreement in the results presented here with other published results. The Spearman rank correlation coefficient ρ ranges from -1 (indicating complete disagreement) to $+1$ (indicating complete agreement).

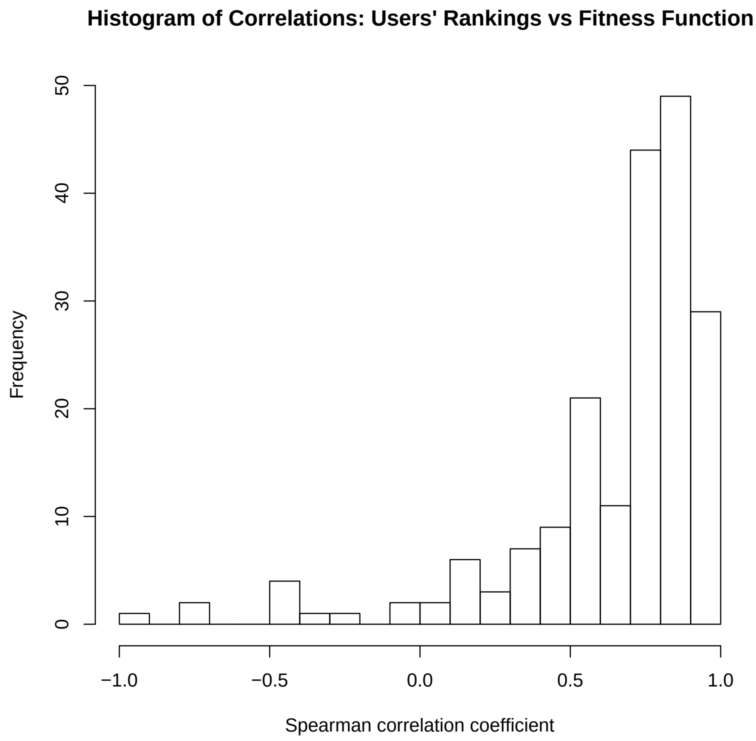


Figure 5.2: The distribution of Spearman rank correlation coefficient scores between rankings of human evaluators and ranking derived from the fitness function score, for all sets of colour schemes.

The correlation between the fitness function and user are all significant at the $p < 0.001$ level with the exception of image set 6. This image set has only a slight and statistically insignificant positive correlation ($\rho = 0.224,\ p = 0.059$).

When image set 6 was examined, the reason for the anomaly was clear. The two highest ranking colour schemes (ranks 5 & 6) were quite acceptable, but the colour schemes in ranks 1 to 4 varied widely and had an evident problem of poor readability, although which item was difficult to read varied. It is not surprising that the human evaluators were inconsistent in their ordering of schemes that were partially unreadable and had unusual and varied colourings.

The image sets for use in the trial were chosen randomly and were not preselected to be “appropriate”. The lack of consistency in the lower ranks of the anomalous image set is consistent with the way that the colour schemes were generated. The dynamics of the genetic optimisation process is such that there can be no guarantee that the intermediate results of the optimisation process will result in schemes with uniform incremental improvement in all four fitness criteria at the same time. It is possible, as happened with this image set, that only the schemes near the end of the optimisation have acceptable values of all four fitness terms.

Image set no.	Spearman rank correlation coefficient		Image set no	Spearman correlation coefficient	
1	0.857	$p < 0.001$	9	0.686	$p < 0.001$
2	0.433	$p < 0.001$	10	0.705	$p < 0.001$
3	0.576	$p < 0.001$	11	0.786	$p < 0.001$
4	0.586	$p < 0.001$	12	0.719	$p < 0.001$
5	0.724	$p < 0.001$	13	0.590	$p < 0.001$
6	0.224	$p = 0.059$	14	0.776	$p < 0.001$
7	0.800	$p < 0.001$	15	0.800	$p < 0.001$
8	0.648	$p < 0.001$	16	0.581	$p < 0.001$

Table 5.5: Correlation (human vs fitness function) for the images within each image set.

Correlations between users and fitness function, by evaluator

To determine the uniformity of the correlations and to find if there were any users who disagreed with the fitness function's assessment, the correlations were reassessed by participant.

Evaluator	Spearman rank correlation coefficient		Evaluator	Spearman rank correlation coefficient	
1	0.104	$p = 0.315$	7	0.564	$p < 0.001$
2	0.504	$p < 0.001$	8	0.740	$p < 0.001$
3	0.714	$p < 0.001$	9	0.732	$p < 0.001$
4	0.779	$p < 0.001$	10	0.818	$p < 0.001$
5	0.768	$p < 0.001$	11	0.564	$p < 0.001$
6	0.721	$p < 0.001$	12	0.860	$p < 0.001$

Table 5.6: Overall Spearman correlation (ρ) between the human rankings of colour schemes within an image set and those of the fitness function, grouped by subject.

As can be seen from the Spearman correlation coefficient scores in table 5.6, eleven of the twelve participants agreed with the rank-order derived from fitness function score. Participant 1 disagreed. It is possible that this participant misunderstood the instructions, or had colour preferences that differed widely from the norm. Nevertheless, the overall level of agreement lends support to the choice of the underlying four terms as being representative of the criteria used in human assessment of GUI colour schemes.

Overall correlation between participants and fitness function rankings

The strong correlation between the rankings of the human participants and the ranking derived from the fitness function score when the correlations are grouped by image set and by evaluator, suggests that they are correlated overall.

The overall correlation between the ranking of the images and the evaluators' rankings (over all images and all evaluators) was found to be Spearman-$\rho = 0.656$, $p < 0.001$. Table 5.7 shows the details of the different correlations, with and without the outliers.

Rankings included	Spearman correlation coefficient
All images rankings by all evaluators	0.656, $p < 0.001$
All image rankings omitting:	
the outlier image set	0.684 $p < 0.001$
the outlier evaluator	0.706 $p < 0.001$
All images, except outlier image set and outlier evaluator	0.731 $p < 0.001$

Table 5.7: The Spearman rank correlations between ranking a colour scheme by human evaluators and the rank derived from the fitness function score.

Image set no.	Kendall-W coefficient of concordance		Image set no.	Kendall-W coefficient of concordance	
1	$W = 0.837$	$p < 0.01$	9	$W = 0.504$	$p < 0.01$
2	$W = 0.202$	$p < 0.05$	10	$W = 0.616$	$p < 0.01$
3	$W = 0.428$	$p < 0.01$	11	$W = 0.733$	$p < 0.01$
4	$W = 0.466$	$p < 0.01$	12	$W = 0.599$	$p < 0.01$
5	$W = 0.602$	$p < 0.01$	13	$W = 0.413$	$p < 0.01$
6	$W = 0.134$	$p > 0.05$	14	$W = 0.660$	$p < 0.01$
7	$W = 0.754$	$p < 0.01$	15	$W = 0.804$	$p < 0.01$
8	$W = 0.527$	$p < 0.01$	16	$W = 0.406$	$p < 0.01$

Table 5.8: The Kendall-W coefficient of concordance assessing the degree of agreement between participants on the ranking of colour schemes within each set. The significance values are derived from a table lookup that does not include values of p less than 0.01.

The results show a strong and statistically significant correlation between the rankings given by the evaluators and those derived from the fitness function score. Before continuing, it is worthwhile determining how closely the subjects agree, not with the fitness function, but with each other in their assessment of the colour schemes.

Determining the consistency of rankings

There are no reference standards for colour scheme appeal or suitability that can be used to assess the "correctness" of an evaluation. However, the experimental data from participants can be used, not only to determine the agreement between the rankings of fitness function and the participants (as shown above), but also to determine how well the participants agreed amongst themselves on the ranking of colour schemes within each set.

The Kendall Coefficient of Concordance known as *Kendall-W* Siegel and Castellan (1988) is a statistical test that measures how closely the rankings from a group of judges match the group mean. This test is used to assess agreement between a number of judges in the absence of an objective standard. The result from Kendall-W is a number between zero and one, where one indicates complete agreement.

The Kendall-W concordance (the agreement) between the evaluators on the rankings of the different images within each of the image sets is shown in table 5.8. The agreement between users on the rankings of images within fourteen of the sixteen image sets had a significance level of $p < 0.01$ – very significant agreement, one $p < 0.05$ and the anomalous scheme, as expected, was higher at $p > 0.05$.

The overall agreement for all images resulted in a coefficient of concordance of Kendall-W$= 0.439, p < 0.01$, again indicating strong agreement. Participant 1, who disagreed with the fitness function ranking also disagreed with the majority view. This individual's coefficient of concordance was almost zero – Kendall-W$= 0.058$, $p > 0.05$. Omitting this outlier from the overall concordance evaluation gave even higher agreement: Kendall-W $= 0.509$, $p < 0.01$. This might be interpreted as a more representative measure of the overall agreement, but, as is indicated by the outlier, not everyone has the same perception of the merit of any particular arrangement of colours. Nevertheless, even when the data set includes the outlier, the overall agreement is highly significant.

Before proceeding to user trials of the generated colour schemes, a further experimental trial was conducted to validate the above results and determine whether the fitness function could be improved.

5.2.3 The main fitness function validation experiment

In the previous experiment, participants ranked colour schemes sampled during the optimisation of a single population of colour schemes. A more rigorous test would use completely unrelated colour schemes within each image set and determine whether a statistically significant correlation still exists.

This experiment is essentially a repeat of the previous colour scheme ranking experiment, but this time using uncorrelated images and a more representative sample of the user population. As a side effect, data from the trial will be used to enable the weights of the four fitness function terms to be assessed.

Removing the dependence between the fitness function colour balance and the wireframe terms

When scoring the colour schemes for this trial, each fitness function term was given equal weight, but there may be non-uniform weights that allow the fitness scores to better match human assessments of the colour schemes.

To allow new values for the weights to be calculated from user assessments of the schemes, it is desirable for the colour balance and wireframe alignment fitness function terms to be independent. However, the presence of an underlying wireframe or path in the colour space is implicit in Munsell's law of colour harmony[6].

[6] e.g. when balancing a large area of a low saturation colour against a smaller area of a higher saturation colour, the colours are expected to be from complementary hues (either side of mid-grey). However, this is not required by the commonly-used formulaic expression of Munsell's law of colour harmony: $a_1 s_1 l_1 = a_2 s_2 l_2$, where a = area, s = saturation and l = lightness.

Non-Munsell colour balance model is used in main fitness validation
Unlike the pilot fitness validation trial, the colour balance term used in this trial is similar to, but not the same as, Munsell's colour harmony theory. The model was modified to balance *area*×*value* and *area*×*saturation* around mid-grey separately (sec. 4.5.2), rather than balancing *area*×*saturation*×*value*, as suggested by Munsell. This is necessary for two reasons. Firstly, separating the lightness and saturation balance makes it possible to use black, white and grey elements freely within the colour scheme – which was found to be necessary in user interface colour schemes – without disturbing the colour balance. For those wireframes that balance around mid-grey, if the developer chooses to allow the "use Black-white axis" option, the lightness-area balance can be disabled without affecting the saturation balance. Secondly, it enables the colour balance term to be assessed without requiring items to be constrained to a wireframe: colour balance assessment is no longer dependent on wireframe alignment. However, altering the colour balance evaluation in this way does mean that the method of creating harmonious colour schemes no longer conforms to Munsell's law of colour harmony, although the modified equations do still incorporate the essential balancing of area against saturation and the less clear-cut (see sec. 2.9.6) area–lightness balance.

The modified colour balance equations with independent colour balance and wireframe alignment terms were used to create the colour schemes for this trial. The ranking data from the participants can be used to refine the weights used during the evaluation of the fitness function. This is described later, in section 5.2.5.

Aim

This experiment has two aims:

1. to determine whether the four term fitness function used in the Colour Harmoniser to assess the "appeal" of an colour scheme is monotonically and positively related to human evaluations of the same colour scheme.

2. to gather the data necessary to determine whether a set of non-uniform weights would result in a fitness function score that more closely matches the human evaluation of the colour schemes than uniform weights.

Experimental design

The design of the experiment is identical to the previous experiment – the participants were asked to rank sets of colour scheme images.

Experimental method

The data set for this trial consists of twelve sets of colour schemes (six complementary, six split-complementary) in which the colours for each of the six schemes within a set were chosen randomly. Therefore, unlike the previous trial, the colour schemes within each image set were completely unrelated.

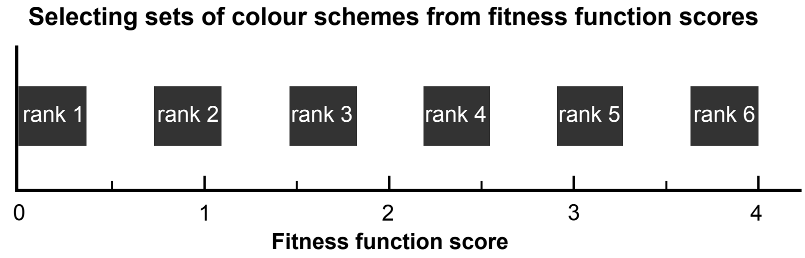


Figure 5.3: The possible range of fitness function values is partitioned in equally size bands and gaps. The colour schemes for $rank_1, rank_2$ etc. were those whose fitness fell in the rank bands. Schemes whose fitness fell in the inter-band gaps were not used.

The mechanics of selecting the random schemes To ensure that the sample images were distributed over the whole range of fitnesses and their fitnesses were sufficiently far apart to produce meaningful distinctions, images were selected from six discrete and well separated fitness bands, as shown in figure 5.3. The gap between the fitness bands ensures that the colour schemes in one rank differ significantly from those in an adjacent band. This allows the schemes to be produced without human selection, to avoid bias. If the schemes were mechanically selected without gaps between the fitness bands, schemes with scores of 2.999 and 3.001 could, if the cutting point between ranks was 3, be associated with differing fitness bands. The enforced gap ensures that the evaluators will not be asked to rank such closely scoring colour schemes.

The widths of the selection bands and the gap were the same. As discussed in section 4.5.1, the fitness function score is not expected to be linearly related to perceived colour scheme quality, but, in the absence of any other basis for choosing the width of the bands and gaps, equal sizing makes the fewest assumptions. Whenever an image score fell within one of the six selection bands, that colour scheme and the associated fitness data, were saved as a sample. Schemes scoring in the inter-band gaps were not used. Each set of colour schemes had one randomly selected scheme from each rank.

The interface elements were coloured randomly and the resulting schemes were scored by the fitness function. When the results were assessed, the mid-level rank bands (ranks 2, 3 and 4) had, by far, the most samples. Finding images with the more extreme fitness values is very unlikely[7]. As will be shown in section 4.21, even with a strong predisposition of the atoms towards the wireframe, only 0.3% of colour schemes score well. In another Monte Carlo search, assessing over 14,000,000 colour schemes (without a wireframe predisposition), no schemes of $rank_1$ (the lowest) or the two highest ranks ($rank_5$ or $rank_6$) were found. Therefore, to find images within these ranks, the optimiser was used and the optimisation terminated whenever a single colour scheme with the required fitness was found. The initial random starting point and the termination after saving a single scheme make it highly likely that the saved schemes

[7] Schemes with a low score require all of the four fitness function terms to score poorly, which is unlikely. A scheme with the atoms well away from the wireframe will have a poor wireframe score but may (if the atoms are widely separated) score well on discriminability. High scoring schemes require all four criteria to score well, which is very unlikely to happen by chance.

will be independent. To generate the $rank_1$ schemes, the optimiser was modified to create low (rather than high) scoring schemes.

As the aim was to assess the coloured elements of the resulting schemes as rigorously as possible, only colour schemes based on the complementary and split-complementary wireframes were used. Monochromatic schemes were excluded and option of including the black-white axis in the complementary/split-complementary schemes was *not* enabled when evaluating colour schemes for this trial.

The colour schemes in each set, being independent, can have elements using completely different colourings (e.g. one scheme using lavender and dark green, another using pale orange and dark blue). It is not clear how the participants would rank two schemes, one of which uses colours they like, the other, colours they would not normally choose, but which are used well. If, from an artistic and readability viewpoint, the unusual scheme was better, would their evaluation reflect this or would their usual colour preferences dominate? Personal colour preference and the use of different colours in the schemes of different rank within each image set may be confounding factors and a source of significant variability.

In informal trials, it was noted that ranking the unrelated schemes was more difficult and took longer than ranking the related schemes in the previous trial. To allow for this, the participants were asked to rank fewer image sets, twelve in this trial rather than the sixteen in the previous experiment.

Subjects and experimental conditions

The nineteen participants were volunteers from staff and students from Massey University, students from the UCOL Design School, members of the public visiting the local TeManawa Science Centre/Museum, and family and friends of the author. The experimental sites included a Design School studio, a museum foyer and private homes. The youngest participant was in the range 15-24 and the oldest 65+, with only five of the 19 participants being teenagers. Unlike the earlier trial, this one was evenly balanced with respect to gender (ten females, eight males, one declined to answer).

The experiment was performed on differing displays, either Phillips 190B 19" home/small office quality or a Phillips 200P 20" semi-professional LCD displays, using an sRGB ICC profile from a GretagMacbeth EyeOnePro spectrophotometer. As in the previous experiment, the lighting was not standardised, but the most common lighting was daylight or incandescent.

Results

The nineteen participants all ranked twelve image sets containing six colour scheme images. To assess the correlation and its significance, the Spearman-ρ correlation coefficients between the human rankings and those derived from fitness function scores were calculated. A histogram of the resulting correlation values is shown in figure 5.4. As can be seen, almost without exception, the correlations are positive.

The overall correlation between participants and the ranking derived from the fitness function was highly significant:

$$\text{Spearman-}\rho = 0.690 \quad p < 0.001$$

with the correlation scores all being positive and, when broken down by image set and by participant (unlike the *Pilot Fitness Function Validation Experiment*), were all statistically significant at the $p < 0.001$ level. The internal agreement between users, evaluated using Kendall-W, was also highly significant:

$$\text{Kendall-W} = 0.5719 \quad p < 0.01^{8}.$$

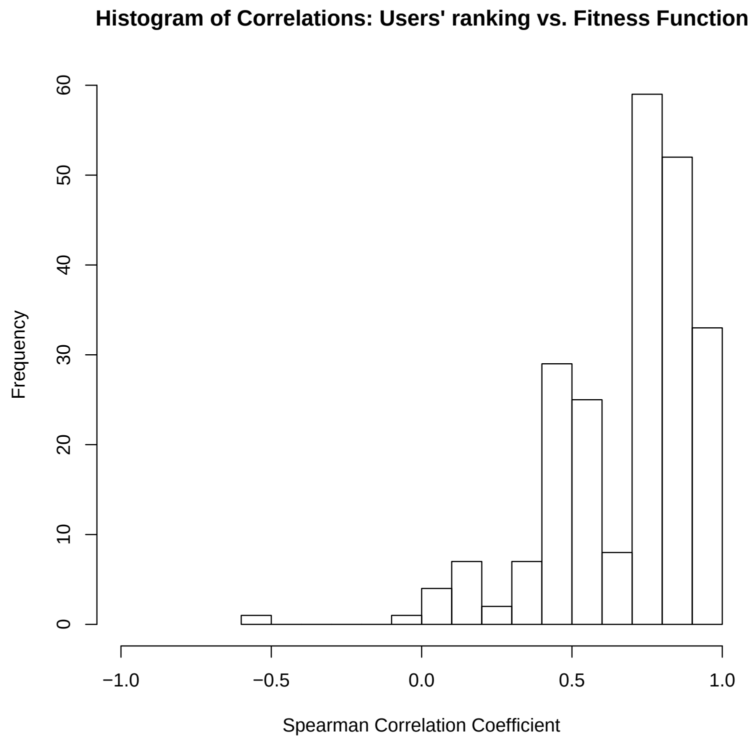


Figure 5.4: The distribution of Spearman correlation coefficient scores (ρ) between rankings of colour schemes having stratified levels of fitness by human evaluators and the ranking derived from the fitness function score, for the all randomly generated colour schemes in data set Main Fitness Function Validation trial.

The high level of agreement confirms the results of the earlier experiment: the scores from the fitness function do serve as a good approximation to the human assessment of a user interface colour scheme.

The fitness function used to score the schemes in this trial combined the contributions from the four fitness function terms using uniform weights. It is possible that uniform weights are not optimal, and different weight values may allow the fitness score

[8] the table lookup method of finding the significance of Kendall-W did not include values less than 0.01.

of a colour scheme to even more closely match human assessment. Therefore, the human rankings of the schemes within this data set can be used to determine whether the uniform weights should be retained, or replaced with experimentally derived weights, and if so, what values should be used.

5.2.4 An experiment to establish the independence of fitness function terms

The fitness of a particular colour scheme is calculated as the weighted sum of four terms (equation 4.21, p136):

$$F = w_i C + w_j W + w_k D + w_l R$$

It is possible that two or more terms in this function are mutually correlated, and that fewer than four terms will suffice.

Aim

To determine whether, in highly scoring colour schemes[9], the four terms of the fitness function are independent.

Experimental design

The degree of independence of each of the four fitness function terms can be determined by measuring the correlations between the scores of each of the four fitness function terms for a statistically significant number of colour schemes with high overall fitness. To generate a large number of highly scoring schemes, the Colour Harmoniser could be run multiple times, or the required schemes could be generated randomly using a Monte Carlo approach. The Monte Carlo method has no underlying colour harmony model and is therefore the least prone to unwarranted assumptions, so it was used to create the required set of colour schemes.

Experimental method

A Monte Carlo data set consisting of a million colour schemes was constructed in which the position of each atom was randomly selected within the abstract colour space, yielding completely random colour schemes. During this process it was found to be necessary to ensure that the atoms were placed in the vicinity of the wireframe, as otherwise the wireframe alignment term would almost always be low[10], penalising the overall score.

To give the randomly selected points a tendency towards the wireframe, but to do so in a random way, the algorithm first placed the atoms randomly within the abstract

[9] the aim is to determine whether all four terms are independent in high-scoring schemes, therefore only high scoring schemes were used for the experiment.

[10] The chance of all the atoms being positioned near to the wireframe is extremely low, but to ensure schemes of a high overall fitness, all the atoms must be close to the wireframe.

colour space and then moved them closer to the nearest wireframe point by a predefined factor (a).

$$p = a \times c(x) \quad + \quad (1 - a) \times x \tag{5.1}$$

where:

p – a point in abstract colour space with an affinity to the wireframe

a – the affinity to wireframe: 0 = none, 1 = on the wireframe, $0 \leq a \leq 1$

x – a randomly chosen point in abstract colour space

$c(x)$ – a function that returns the closest point on the wireframe to x

The raw data set consisted of one million schemes using a split-complementary wireframe, with the wireframe affinity (a in equation 5.1) set to 0.9. From this raw data set, colour schemes with high fitness were selected for use in the correlation analysis.

Results

Of these one million schemes, even with the wireframe affinity, only 3109 (0.3%) were of rank five or six, as is shown in the histogram of the fitness function evaluation of the resultant schemes in figure 5.5.
To determine if there was any visually evident correlation between the terms, a scatter plot of the 3109 higher ranking schemes was produced (fig. 5.6). The terms appear mostly uncorrelated with the exception of *colour balance:readability* and *colour balance:distinguishability*, which have a negative slope. The Pearson correlation coefficient (Table 5.9) quantifies the visually apparent correlation: readability is more strongly correlated (-0.605) with colour balance than distinguishability (-0.365). The

	Colour Balance	Wireframe Alignment	Distinguish-ability	Readability
Colour Balance	1.000	-0.125	-0.365	-0.605
Wireframe Alignment	-0.125	1.000	0.052	-0.059
Distinguishability	-0.365	0.052	1.000	-0.176
Readability	-0.605	-0.059	-0.176	1.000

Table 5.9: The correlations between the four fitness function terms for randomly generated schemes with high overall fitness (method: Pearson's product moment correlation)

negative correlation between colour balance and readability (and to a lesser extent distinguishability) is to be expected. The interface used in the trial has one dominant interface item: the body background that accounts for over half the area. For the colour strength balance score to be high when using a complementary or split-complementary wireframe, several smaller items would need be on the opposite side of the wireframe to the body background to act as a counterbalance. This would place both a textual item and its background (e.g. button text and button background) onto same side of the wireframe, reducing their lightness difference and therefore decreasing the readability score of the text. This increases the score of one term while decreasing the other, result-

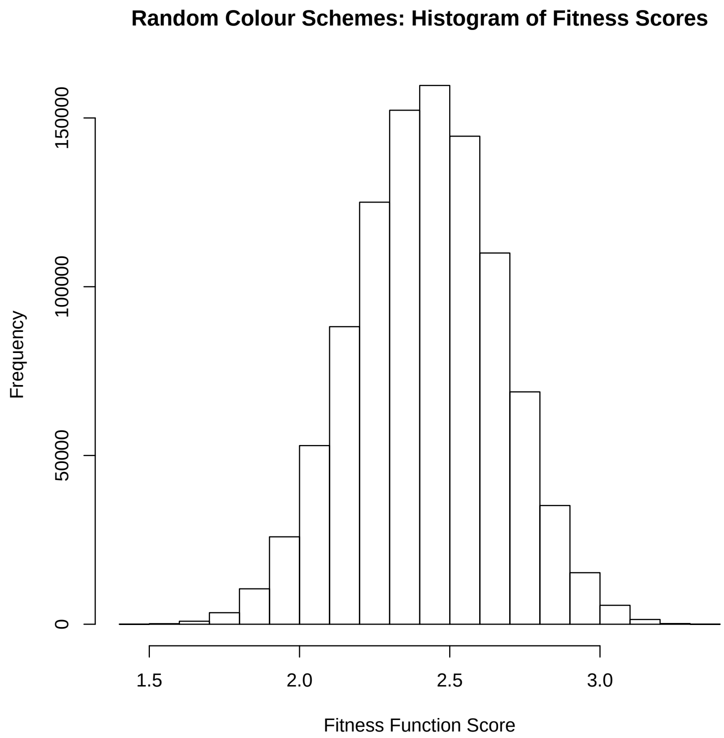


Figure 5.5: The histogram of fitness function scores of one million colour scheme generated by a Monte Carlo method, using a split-complementary wireframe, with a wireframe affinity of 0.9. Even with such a high wireframe affinity, very few (~0.3%) acceptable interface colour schemes result from random colourings.

ing in the observed negative correlation. The lower correlation for distinguishability is reasonable: distinguishability is less constrained as any difference in colour space (ΔE) is acceptable, not just vertical (lightness) difference (ΔL). Therefore, items on the diverging arms of the Y-shaped split-complementary wireframe could give acceptable distinguishability scores, but not satisfy the readability requirement.
There is no unexpected correlation between the four fitness function terms and so, to ensure the initial reason for the inclusion of each term is taken into account during the optimisation of the colour schemes, all four terms must be included in the overall fitness function.

5.2.5 Estimating the fitness function terms weights from experimental data

The data is plotted (fig. 5.7) to determine whether there is any visual evidence of a relationship between the fitness function scores and the ranks given to images by the users. The box plot[11] shows the range of fitness function scores for the images assigned

[11] Interpretation of the "box and whisker" plot: this plot is used to show the summary statistics of a data set, such as its median, its distribution, and its range. The box extends from the 1st to 3rd quartiles (25% to 75%) – 50% of the data values are within this range. The median is shown by the

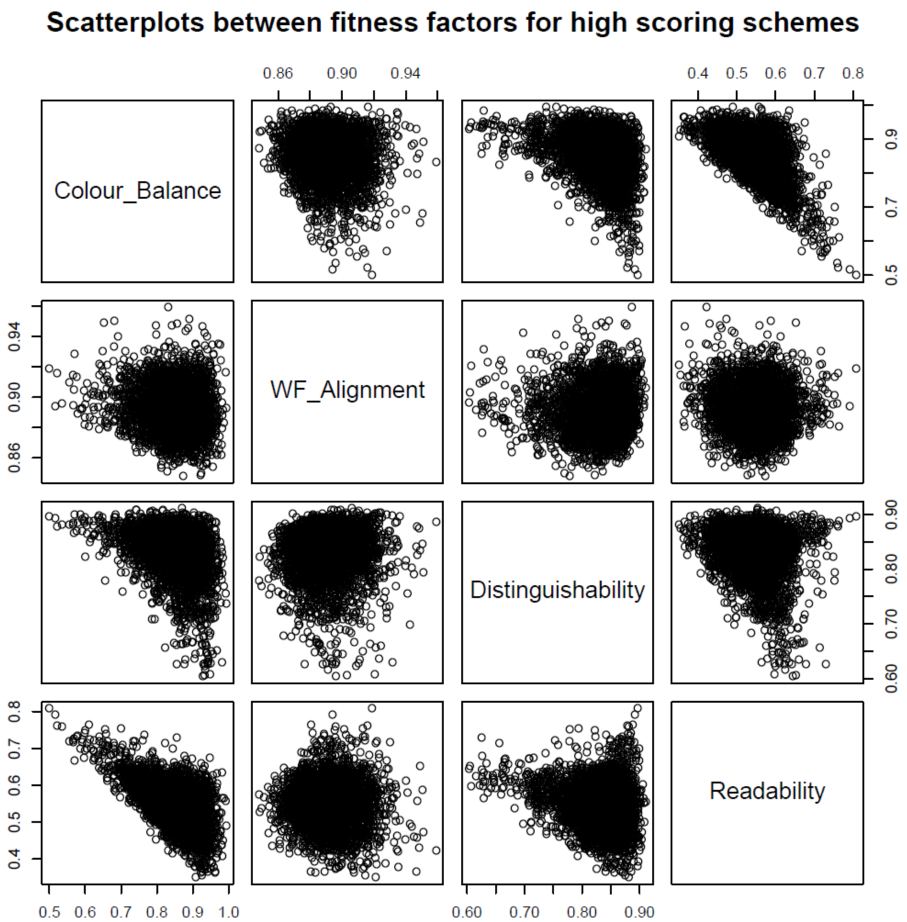


Figure 5.6: A scatterplot between the scores of each fitness function terms for randomly generated schemes having a high fitness function score. To get high scoring schemes, a forced wireframe affinity was incorporated into the random scheme generation.

to each rank by the evaluators.

For the schemes in $rank_2$ to $rank_6$, there is a clear upward slope between the scores assigned by the fitness function to colour schemes and the scheme's ranking by participants. The higher median value in the lowest rank is a little surprising, indicating that a few of the high scoring schemes were not liked by users. This is not surprising as schemes using some hues are not particularly pleasing even when the inter-element relationships are as expected. Such schemes at another hue rotation (after human adjustment) can be quite acceptable. More importantly, there were hardly any low scoring schemes that were highly ranked. A large number would indicate something missing from the fitness function. However, a small number of pleasing, but low scoring, schemes is acceptable, and could be interpreted as meaning that, even amongst schemes

darker line within the box. The whiskers (dotted lines) show the limits of data values to a maximum of 1.5 $\times$ the interquartile range (the length of the box). Values beyond the whiskers are considered outliers and are shown by the little circles, the limits of which (if present) show the range of the data.

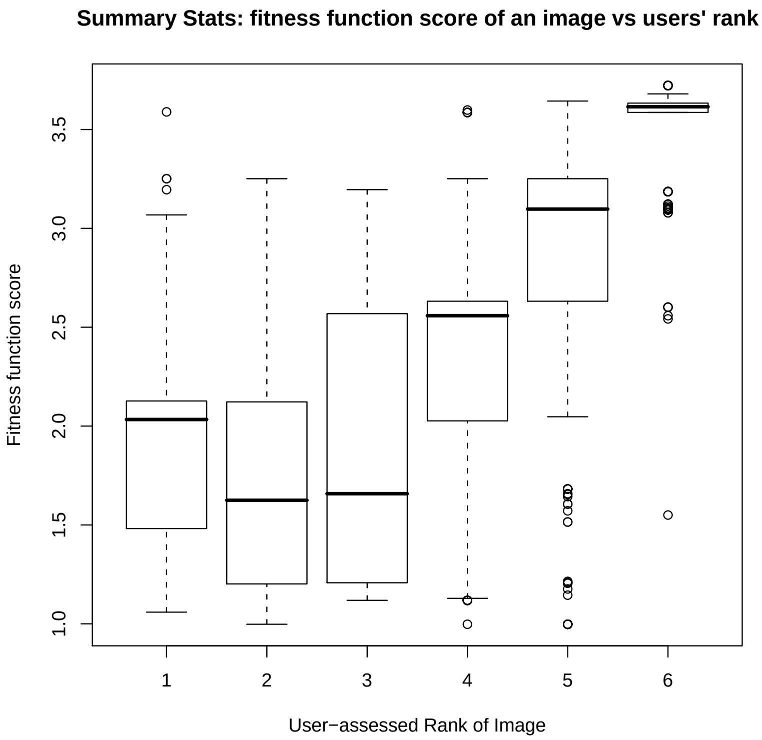


Figure 5.7: The summary statistics for the fitness function scores plotted against user-assigned image rank for the colour schemes in data set *Main Fitness Function Validation Experiment*. The darker line inside each box indicates the median fitness score and outlier values are shown by the small circles.

that are quite different from those normally classed as good, there is the occasional one that some people find attractive. Alternatively, of course, the fitness function may need some improvement. This is visual confirmation of the previously noted correlation between the ranks determined from the fitness function score and the ranks assigned by users:

$$\text{Spearman}-\ \rho = 0.690, \quad p < 0.001$$

Given that the participants agreed with the ranking of images determined from their fitness function score, and the visually evident relationship in fig. 5.7, the data from the user evaluations is a solid basis for estimating a revised set of weights.

Are all four fitness terms equally important?

The lack of any strong correlation between the four fitness function terms indicates their independence, but before attempting to evaluate the appropriate weights for each term, it is reasonable to ask whether all four terms are actually necessary. While it is convenient to think of each of the four as being (more or less) uniformly important, this not necessarily so. This can be understood by considering the effect of omitting each term, a summary of which is shown in Table 5.10.

The colour strength balance and wireframe alignment terms are included as part of the codification of the artistic model of colour harmony and will therefore be included. If the readability term is omitted, the optimiser can (and does) return high scores for schemes in which one or more elements are unreadable, and therefore unusable, regardless of how prettily the colours are used. The readability term is therefore essential.

Of the four terms listed above, only the distinguishability term could be considered optional. If it is omitted, the resulting schemes will still be harmonious and readable, but areas that should be coloured differently may not be. For many interfaces, this would only affect the aesthetics. Therefore the evaluation of optimised weights will proceed without the inclusion of a distinguishability term – a constraint that can be set (or not) at the whim of the developer. The distinguishability term is not going to be omitted from colour scheme optimisation, just from the evaluations of weights.

Fitness Function Term	Effect of Omission
Colour balance	• saturation and light-dark balance will be affected. • large areas of strongly saturated colours are possible.
Wireframe Alignment	• the progression apparent when colours lie on a line or a curved path in the colour space will be missing. • interface element colours will be scattered throughout the colour space, potentially resulting in discordant colour schemes.
Distinguishability	• separate items (*other than text*) may be coloured identically or so closely so as to be indistinguishable. Text, having its own fitness function term, is treated as a special case and will always be distinct from its background as long as the *Readability* term is operative. • user-specific aesthetics preferences are not enforced. • items depending on colour distinction for their semantics may lose their meaning.
Readability	• one or more items of text may become difficult to read or disappear altogether.

Table 5.10: The effect on colour schemes resulting from the omission of each fitness function term.

Finding the revised fitness function weight values

If the users had evaluated the fitness of each image on a scale of zero to four, it would be possible to perform a multiple regression of the user evaluations as a function of each of the fitness terms to obtain an optimum set of weights.

The primary aim of main fitness function validation experiment was to determine whether the fitness function evaluations agree with those of human evaluators, and for this, ranking was used, for the reasons discussed in section 5.2.2. Therefore, the data available from the experiment are ranks not scores. However, the user-assigned ranks of the colour schemes can used as an approximation to the user-assigned scores. The user-assigned ranks 2 to 6 and the fitness function scores have a strong positive correlation (Pearson correlation coefficient = 0.73) and they have the same numeric range (2 to 6 for the user ranks vs. 0 to 4 for the fitness function scores). Over the 1368 samples[12], the user-assigned rank could be expected to be similar to the values

[12] 1368 rank values, from 19 users × 12 sets of colour schemes × 6 images per set.

that would have been assigned by scoring the schemes on a discrete scale.

Therefore, the evaluation of weights can be determined by performing a linear regression of the terms being combined to make up the fitness function score for a colour scheme against the rank (acting as a score approximation) of that scheme assigned by human evaluators. In other words, finding a set of weights that minimises:

$$\sum_{j=1}^{1368} \left| r_j - \sum_{i=1}^{3} (w_i . f_{i,j}) \right| \tag{5.2}$$

where:

j – an index that identifies a particular colour scheme

i – an index that identifies a fitness term,

1 = colour strength balance, 2 = wireframe alignment, 3 = readability

r_j – the user-assigned rank, used as a user score approximation, for colour scheme j

w_i – the weight applied to fitness function term i

$f_{i,j}$ – fitness score for fitness function term$_i$ for scheme$_j$

The ranks given by the human evaluators range from two to six, so the estimate calculated would yield a fitness function score outside the desired range of zero to four. However, only the ratios between the weights of fitness function terms are relevant, so the difference in range can be adjusted by rescaling the estimated weights while maintaining their ratios. The result from a regression of all colour schemes ranked by users except those of the lowest rank (i.e. rank ≥ 2) against the score of each fitness function term for that scheme is shown in table 5.11. Normalising these weights so they have a total of four (for comparison with the uniform weights) gives new weight values shown in table 5.12.

	Estimate	Std. Error	t value	Pr(>|t|)
Colour strength balance	1.2171	0.1928	6.3137	p<0.001 ***
Wireframe alignment	2.2838	0.1477	15.4675	p<0.001 ***
Readability	3.1896	0.1668	19.1260	p<0.001 ***

The significance: p: 0 <***<0.001<**<0.01<*<0.05<.<0.1

Regression $R^2 = 0.92 \quad p < 0.001$

Table 5.11: The estimated term weights calculated from a regression of fitness function term scores against the user-assessed ranks two to six.

The distinguishability term, while omitted in the estimation of fitness function weights, must be used in the optimisation, otherwise the user-defined distinguishability preferences would have no effect. To allow for this, the weight given to distinguishability is set equal to the lowest weight of the other three components. As altering the weights of the four terms will result in a different search during the optimisation, rather than attempt to assess what the rankings with the new weights might have been, the new weights (shown in table 5.12) were tested by using the new weights to create the colour

Fitness function term	Optimised Weight
Colour strength balance	0.6157
Wireframe alignment	1.1552
Distinguishability*	0.6157
Readability	1.6134

Table 5.12: Optimised fitness function weights estimated from colour schemes ranked by evaluators at rank 2 and greater. Note: distinguishability weight is not determined from the regression, but is set equal to the lowest of the other three terms.

schemes that are assessed by users during experimental trials. The results of these trials are discussed in the next chapter.

The ranking experiments show that the fitness function's evaluation of the appeal of a user interface colour scheme is in good agreement with the evaluation produced by an average human assessor.

Two terms, colour balance and wireframe alignment, contribute directly to the assessment of colour harmony. Of the two, wireframe alignment has the much higher weight and is therefore more important. As wireframe alignment is the method of choosing an ordered set of colours to be used in a scheme, the higher wireframe alignment weight can be seen as confirmation of the preeminent place of order in colour scheme design, as discussed in section 2.9.2.

5.3 Conclusion

This validation of the fitness function, together with the architectural details contained in the previous chapter, concludes the overview of development of the trial implementation of a Colour Harmoniser prototype. As demonstrated by the screen-shots of the colour schemes created by the Colour Harmoniser prototype in Chapter 4, it is indeed possible to implement the envisaged system.

The Colour Harmoniser prototype is a realisation of the architecture outlined in Chapter 3. The validity of the fitness function has been confirmed experimentally, and that function together with the improved set of term weights will allow the creation of optimised interface colour schemes. These schemes can now be evaluated to determine how they compare to human-created colour schemes.

Chapter 6

The experimental evaluation of the Colour Harmoniser

The approach to colour harmony that has been presented in this thesis extends conventional algorithmic approaches to colour harmony to include considerations pertinent to user-interface design. The previous chapter described the Colour Harmoniser, a software tool that was implemented following this approach, and some experiments that were designed to optimise the values of a number of parameters used by the Colour Harmoniser to direct its search for colour schemes (sec. 5.2). The results of these experiments demonstrate that there is a strong correlation between the fitness function and human assessment of colour schemes. Encouraging though these results are, they do not demonstrate that the Colour Harmoniser method produces pleasing and usable interface colour schemes.

Accordingly, a number of experiments involving user trials, with a focus on comparing the quality and usability of this approach and conventional one-colour-at-a-time approaches, were undertaken. The first "Compare The Methods" experiment involved creating and evaluating colour schemes, whereas the final "Compare The Results" experiment involved only the evaluation of colour schemes.

Compare the Methods: the purpose of this experiment was to gain some insight into the relative difficulty of using the Colour Harmoniser – as a prototype embodiment of the approach to colour harmony – when compared to conventional approaches to colour scheme design. Each participant prepared two colour schemes: one using the Colour Harmoniser method and the other using a traditional method of picking colours for interface elements individually. When both tasks had been completed, the ease of development and the quality of the resulting schemes were assessed and compared.

Compare the Results: the purpose of this experiment was to gain some insight into the quality of the colour schemes produced by the Colour Harmoniser method. Each trial participant evaluated multiple colour schemes without being aware of how the schemes were created. Comparing the resulting evaluation scores enabled colour schemes created using the Colour Harmoniser method to be compared with schemes created by human designers (without constraint), and with randomly coloured schemes.

6.1 Compare the Methods: traditional vs. Harmoniser-based scheme creation

6.1.1 Aims of the experiment

This experiment had three aims:

1. to compare the usability of the Colour Harmoniser method of selecting and optimising colour schemes with the usability of the conventional "one-at-a-time" (manual) method.
2. to allow the participants to evaluate the quality of the colour schemes resulting from both methods.
3. to compare the time taken for participants to create colour schemes using both methods.

A side effect of this experiment was a set of Harmoniser-based schemes that had been adjusted (personalised) by human participants. These scheme along with manually created schemes are needed for the final evaluation experiment.

6.1.2 Experimental design

The experiment used a within-subjects design, in which each participant assessed the quality of the colour schemes she or he had produced, and the ease of producing them, for an identical interface coloured using both the Colour Harmoniser and a conventional approach. The experiment used the comparison methodology, with the same user interface being coloured by the same person using two different methods. The participants were asked to evaluate the quality of the colour schemes and ease-of-use factors for the two different methods separately, and comparatively. Demographic data, including details of the subjects' previous colour design experience and factors relating to artistic self-perception were also collected.

Participants were volunteers, and while there was no explicit matching of age, gender, artist or not, or colour design background, the sites chosen for data collection ensured that there were likely to be a balanced mix.

6.1.3 Experimental method

A software interface was designed to allow the selection and personalisation of Harmoniser-created schemes, and a separate interface to allow the manual colouring of individual interface elements. Each participant created one colour scheme using each method, the presentation order being chosen randomly. Each trial progressed through the following stages:

1. *an introduction:* This outlined the aims of the experiment, what the participant was being asked to do, the details of who was running the experiment, and their organisational affiliation.

2. *a demographic survey:* participants provided their gender, age, artistic-self-perception and colour training. Each question had a pop-up hint expanding (where appropriate) on the question, or giving the rationale for its inclusion.

3. *the creation of two colour schemes:* There were two tasks, each followed by a questionnaire with questions related to the difficulty of the task. For each participant, the order of the tasks was chosen randomly:

   (a) Each participant created a colour scheme for the test interface using a conventional "one-at-a-time" colour selection method. The interface was designed to minimise the user interaction required to select a colour, so that the Colour Harmoniser was being compared with an interface that was easy to use.

   (b) Each participant optimised a Harmoniser-created scheme. The mechanism will be detailed shortly, but in outline:

   - a number of Harmoniser-created schemes were presented to the user
   - the user selects one of these
   - the user can then modify the colouring using a direct-manipulation interface to alter:
     - the colours used (the hue rotation)
     - the intensity of the colours (the saturation control)
     - which colours are light and which are dark (the "flip light/dark" control)

   The hue and saturation controls are continuous, the light/dark control is binary. The interface is non-modal, with the colours in the interface being updated immediately.

4. *a post-creation questionnaire:* there was a separate questionnaire displayed after a colour scheme had been created manually, and using the Harmoniser-based method. They included questions relating to the usability of each method and the perceived quality of the resulting scheme, and the colour scheme was displayed on the questionnaire page, to avoid memory effects[1] or the need to return to the previous page.

5. *a direct comparison of the resulting colour schemes:* both of the colour schemes that the participant had created were shown side-by-side and the participant was asked to select the one that seemed more professional.

[1] Trying to remember a scheme's appearance could be a source of error, as colour memory can be poor (Sawahata (1999) citing Albers (1963))

6. *a concluding survey:* this questionnaire contained questions asking the participant to compare usability of both methods (unlike the post-creation questionnaires above, whose questions related to a single method of colouring).

No time or iteration limit was placed on users, but these details were recorded. No information that could have been used to identify participants was stored with the data.

It was desirable for both the colourings and the comparisons to be completed in a single sitting. This avoided the need to identify returning participants, to handle incomplete trials (from those who did not reappear, which was likely, as the participants were all unpaid volunteers) and it avoided possible memory effects. Consequently, significant effort was expended to simplify the experimental process and so shorten the time necessary, and to minimize any need for training or explanation.

Data needed to run the experiment The manual creation of colour schemes by picking individual element colours does not require any pre-existing data. However, before the participants can optimise a Harmoniser-based scheme, such a scheme must already exist, as the creation of the scheme itself with the requisite explanation of groups and constraints is not part of the trial. Therefore, eighty high-scoring[2] colour schemes were selected from those created by the Colour Harmoniser prototype. The user-definable values were set to the same configuration as in the fitness function validation experiments described in the previous chapter:

- *element grouping:* the following were each put into their own group: header and footer text; header and footer backgrounds; the text on all the buttons; and the background colours of the buttons.

- *distinguishability:* the following were all set to be mutually distinguishable: body background colour; header/footer background group; button background group; the left hand navigation bar background; the colours of all textual items were left unconstrained.

- *wireframe tilt:* consistent with all the schemes generated previously, the wireframe tilt angle was 45° to the horizontal.

- *the LAB Scaling factor:* the value used was 56, as discussed in section 4.7.2.

- *hue rotation angle:* this was chosen randomly to ensure a variety of colours in the resultant schemes.

- *wireframe shape:* a total of eighty schemes were selected, twenty from each of the complementary, split-complementary, analogous and monochromatic wireframe shapes. For the twenty complementary and twenty split-complementary, ten of each included the use of the black-white axis and ten did not.

[2] fitness values of ≥ 3.58 out of 4, which are the same range of values used when selecting the high fitness ($rank_6$) images used in the second fitness function evaluation experiment (sec. 5.2.3).

- *fitness function weights:* the optimised weights detailed in section 5.2.5 were used.

The eighty schemes were themselves randomly selected from a larger pool. Of the eighty schemes, sixty-five were distinct. The remaining fifteen schemes were based on several different colour molecules with differing wireframe rotations, resulting in different colour schemes. To avoid preselection bias, a different set of ten schemes was randomly selected from the eighty as “starter” schemes for each participant (see fig. 6.7, p207).

6.1.4 The interfaces used to control the experiment

The mechanics of conducting the experiments are complicated by the unfamiliar nature of the task, and so software was created to administer the trial and save the resulting data. A strong emphasis was placed on simplicity and transparency. The colouring task was made as uncomplicated as possible:

- the visual complexity was reduced by removing all unnecessary items from the screen;
- simultaneous contrast effects were minimised by presenting the test interface on a mid-grey background;
- the amount of instructional text was minimised;
- the survey questions were simplified, but with additional detail available from popup hints if required.

Designing an experiment that allows the user to focus on the colouring task, without being overwhelmed by irrelevant detail (e.g. selection methods and navigation) presented significant difficulties. The two methods of colouring present distinctly different problems, so two custom interfaces were created. The interface that was designed to allow the easy creation of manual colour scheme will be discussed first, followed by a discussion of the design of the interfaces created to allow the optimisation of a Harmoniser-created colour scheme. Both are illustrated by screen-shots of the stages during each colouring to allow the experimental tasks to be understood. In practice, the participants found the interfaces easy to work with and required little in the way of additional explanation.

It was necessary during the experimental trial to refer back to the methods that were used to create a colour scheme. Rather than use the terms “manual” and ”Harmoniser” (or any variation thereof), terms more appropriate to the experimental context were used instead: *One-colour-at-a-time* was consistently used to refer to the manual colouring of individual interface elements, and *All-colours-together* to refer to the holistic adjustment of the Harmoniser-based colour scheme.

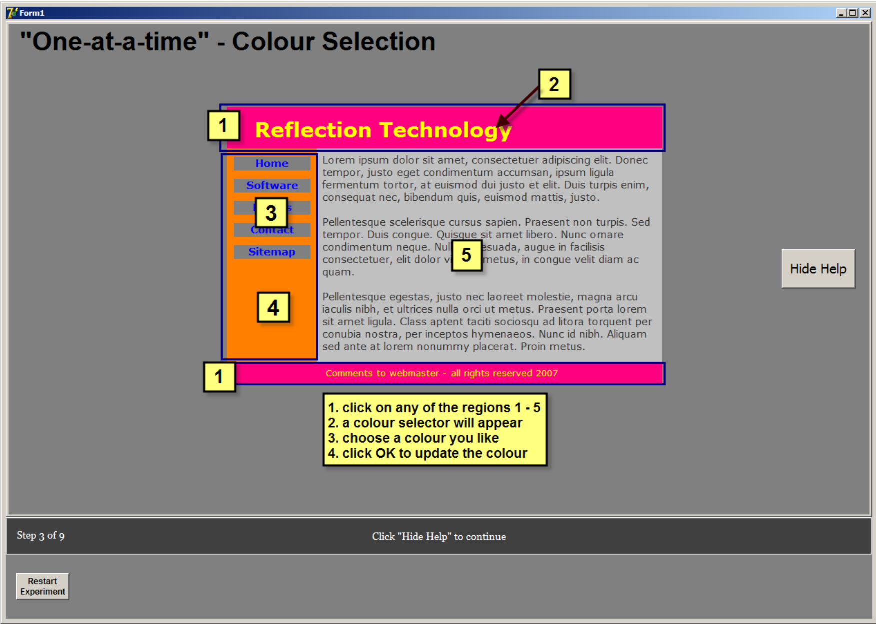


Figure 6.1: Manual colouring trial: this brightly coloured instruction page to the "one-colour-at-a-time" experimental interface shows the colourable regions of the web page, and the simple four-step instructions.

An interface to allow unskilled users to create a colour scheme, selecting colours individually.

Conceptually, it should not be difficult for users who are familiar with computers to choose the colours for seven interface elements. In practice, however, some of those familiar with computers may have little experience with colour-selection tools, and even those who do have such experience will have used a variety of interface types. The conventional method of changing the colour of an interface elements is indirect, and puzzles many unsophisticated computer users.

It is first necessary to select an object and then somehow indicate what is supposed to happen to it, but there are many methods used to accomplish this. The methods used to select multiple objects vary as does the forms of visual feedback to show what is selected. Deselection methods are inconsistent, and the methods to invoke a colour dialog vary. As no interface can be expected to be unfamiliar to all, these points would all require explanation, which would lessen the likelihood of participation and completion.

As there is no guarantee of a common culture amongst the participants, an interface was designed that was be sufficiently simple that virtually no explanation was required and placed experts and inexperienced users on a more equal footing. The user interactions are minimal: the user simply clicks on an item[3] to display the colour selection dialog so a new colour for that item can be chosen. This is less obtuse than the usual

[3] The custom interface also embodies the preset "same-coloured" groupings for buttons, button text, header/footer text and background: setting the colour of one button propagates the change to all the other buttons, and changing the header text colour sets the footer text colour, etc.

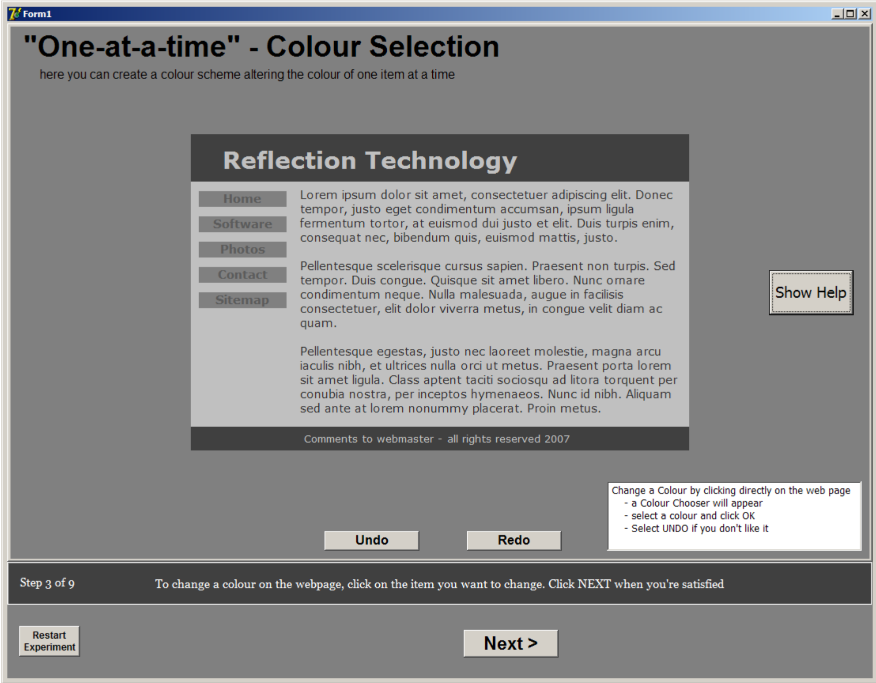


Figure 6.2: Manual colouring trial: the test interface as initially seen is achromatic. Clicking on any web page element will cause a colour selector to appear. The "Show Help" will cause the previous instruction page to reappear.

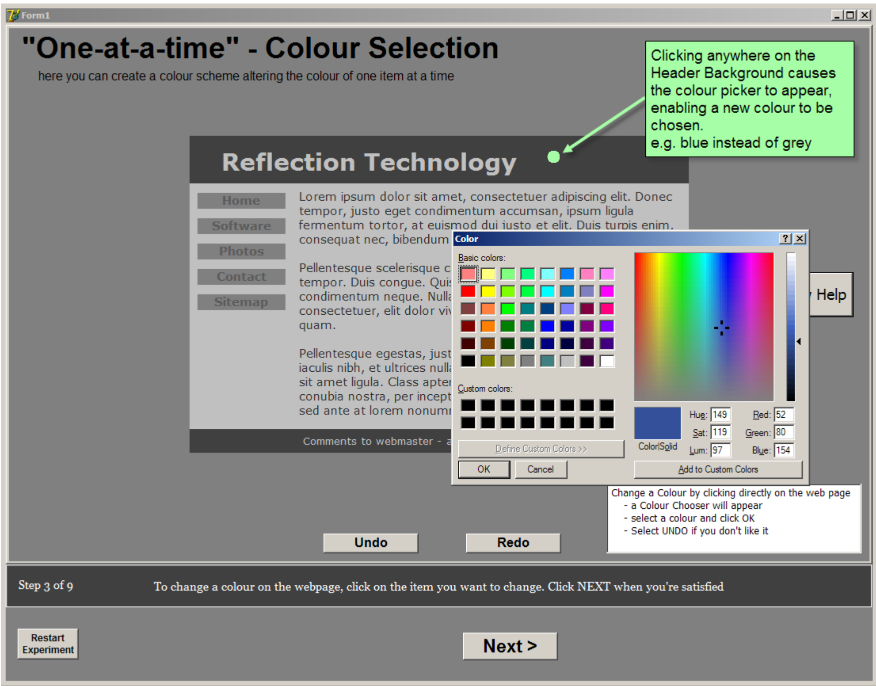


Figure 6.3: Manual colouring trial: clicking on a page element (the header is shown here) displays the standard *Windows*^{TM} colour selection dialog, expanded to show the RGB and HSL options, not just the basic colours. Whatever colour is selected will become the new colour of the clicked-upon element.

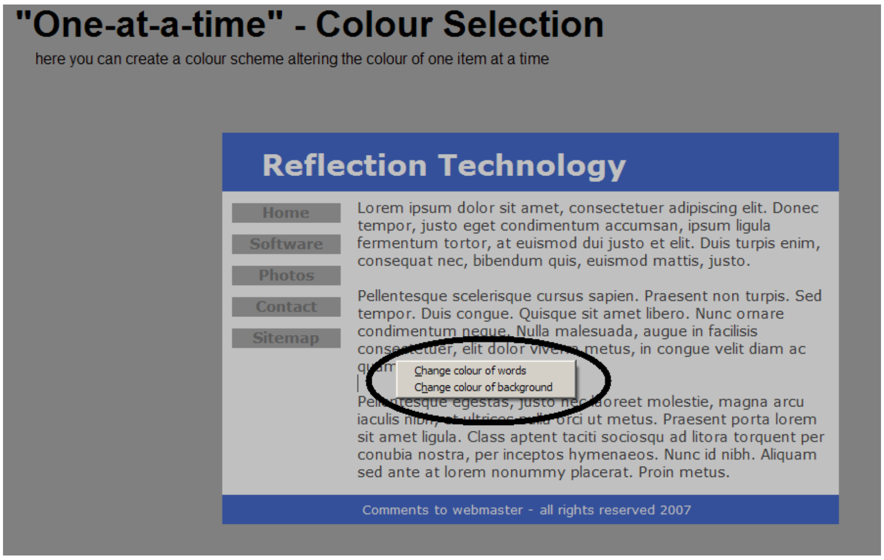


Figure 6.4: Manual colouring trial: the body text and button text are difficult to click on precisely. Therefore, for these elements, the interaction has an additional step: clicking anywhere on body or on a button will cause a popup-menu to appear. When either the text or background is chosen from this menu, the colour dialog is displayed. The additional step ensures the item to be recoloured is the intended one.

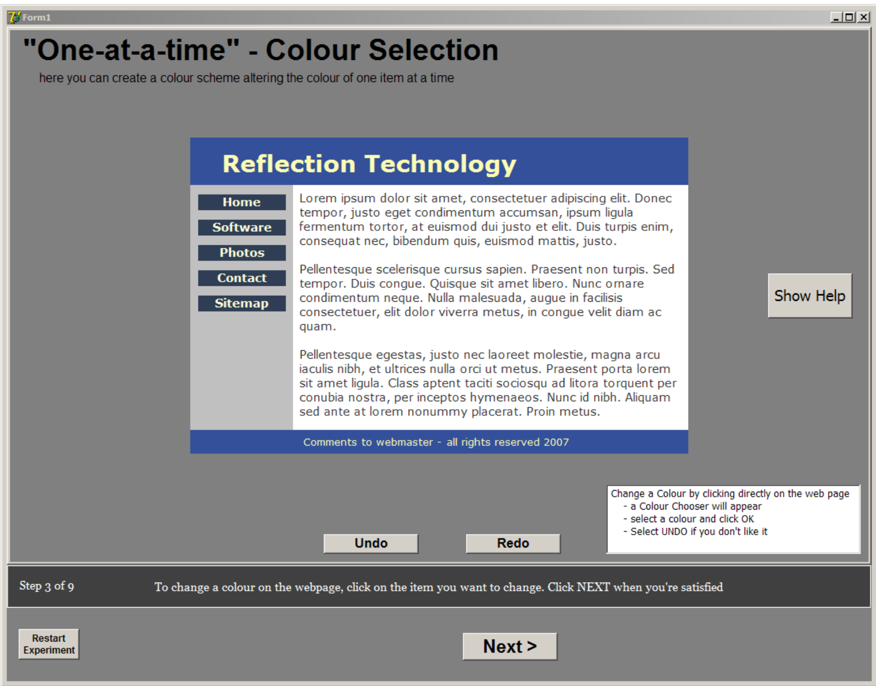


Figure 6.5: Manual colouring trial: the participant can change element colours as often as desired until a satisfactory scheme has been found. Clicking “Next” will move to the post-colouring questionnaire.

multi-step method.

A single demonstration was sufficient to teach the participants how to use the interface, and its uncluttered appearance allows a user to focus on the colour scheme without the distractions of navigating menus and toolbar buttons. The uncluttered structure of the interface also allows it to be positioned in the centre of a mid-grey background, away from other items, which minimized simultaneous contrast effects caused by items external to the web page.

The experimental interfaces seen by a participant when creating a colour scheme manually during an experimental trial are shown in figures 6.1 to 6.5. After the participant had finished creating a scheme, the post-colouring questionnaire was displayed. This will be discussed later, together with the post-colouring questionnaire from the Harmoniser-based colouring method (in section 6.1.4).

An interface to allow unskilled users to optimise a Harmoniser-based colour scheme.

This part of the experiment was intended to ascertain whether a user can select and adjust a Harmoniser-based colour scheme to give a result they liked, to measure the time taken, and to survey their feelings on the quality of result and usability of the method. It was not intended to assess the Colour Harmoniser interface that is used for the interface characterisation phase (capturing groups and defining constraints) and for producing the raw schemes. It is intended to assess the effectiveness of the personalisation phase: the choice and adjustment of raw Harmoniser-created schemes, and to assess the quality of the schemes after personalisation. Therefore, the existence of a pool of Harmoniser-created schemes that have not been subjected to human fine-tuning is assumed. From this pool, a set of schemes is selected and presented as candidate schemes for the user to fine-tune.

It is not straightforward to display multiple colour schemes, any of which the user can modify, without the process being confusing the viewer or requiring training. This method of colouring is probably unlike any method of colouring that the participants have used previously, so, as was done for the manual colour method, a custom interface was designed to minimise complexity. After several iterations, a pair of interfaces was found that would satisfy these requirements. The result is a pair of interfaces, the primary interface is the "home" screen for the Harmoniser-based colouring trial, with the second interface (the holistic colouring adjustment interface) begin invoked when required.

The primary interface:

- displays a variety of potential "starter" schemes created by the Colour Harmoniser prototype.
- has a simple method for the user to select a scheme to recolour (edit).
- shows all the currently available schemes, including any changes made by the user to the initially-supplied Harmoniser-created schemes.

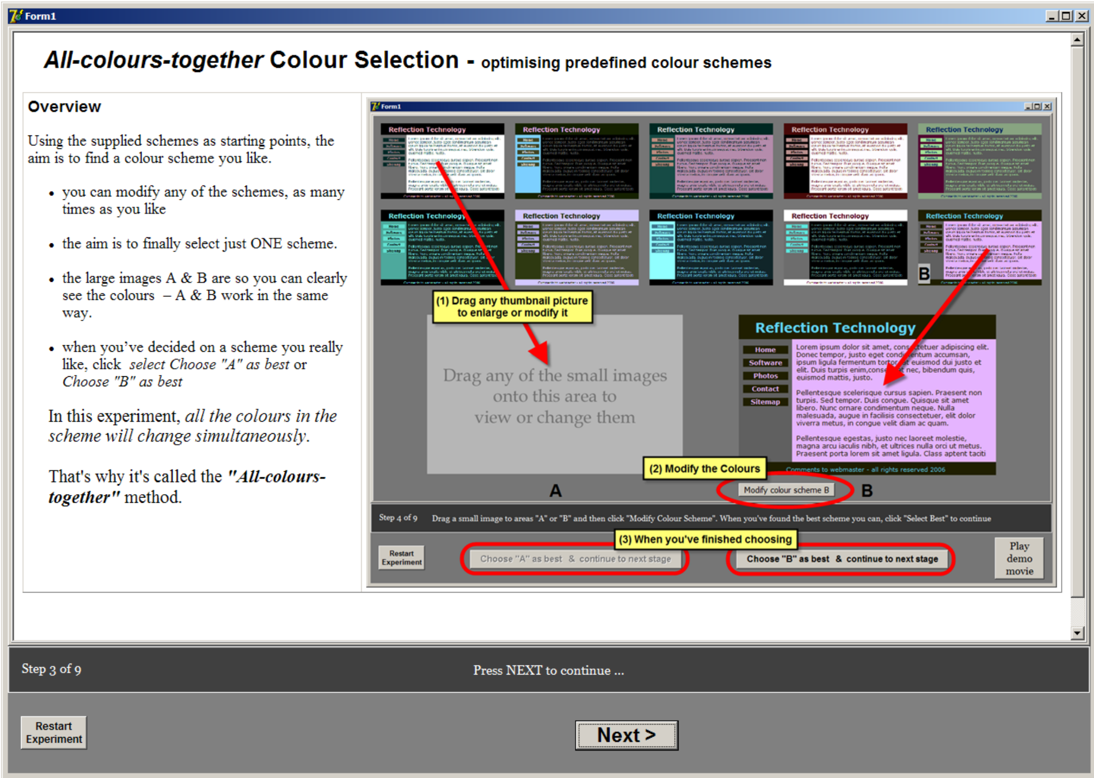


Figure 6.6: Harmoniser-based colouring: the introductory page showing the experimental interface the participants used to choose Harmoniser-generated schemes and invoke the adjustment interface.

- enables any pair of schemes to be enlarged from thumbnails to show the colours more clearly. This enables the final selection of the preferred scheme to be made by direct comparison, rather than from schemes viewed alternately.
- allows (using a second interface) the colour schemes to be adjusted holistically, using the hue rotation, saturation, and "flip-light-dark" controls.
- unambiguously shows the current state and what options are available

The first page that relates to the Harmoniser-based colouring (fig. 6.6) shows thumbnails of ten Harmoniser-created schemes.

Any pair of schemes can be selected and enlarged, allowing the subtle differences in the colouring to be seen, but more than two is not practical due to the limited screen area. Either of the selected pair may be modified, which causes the appearance of the second "colour-change" interface (fig. 6.9). Upon return from the "colour-change" interface, the colours of the selected image and its thumbnail are updated.

The user can drag any of the thumbnails to one of the "A" or "B" enlarged images, either for continued editing, or to select that image (either A or B) as their "chosen" colour scheme. There is no loss of colour scheme information when dragging a new thumbnail to replace an existing scheme in either of the lower A/B windows, as the

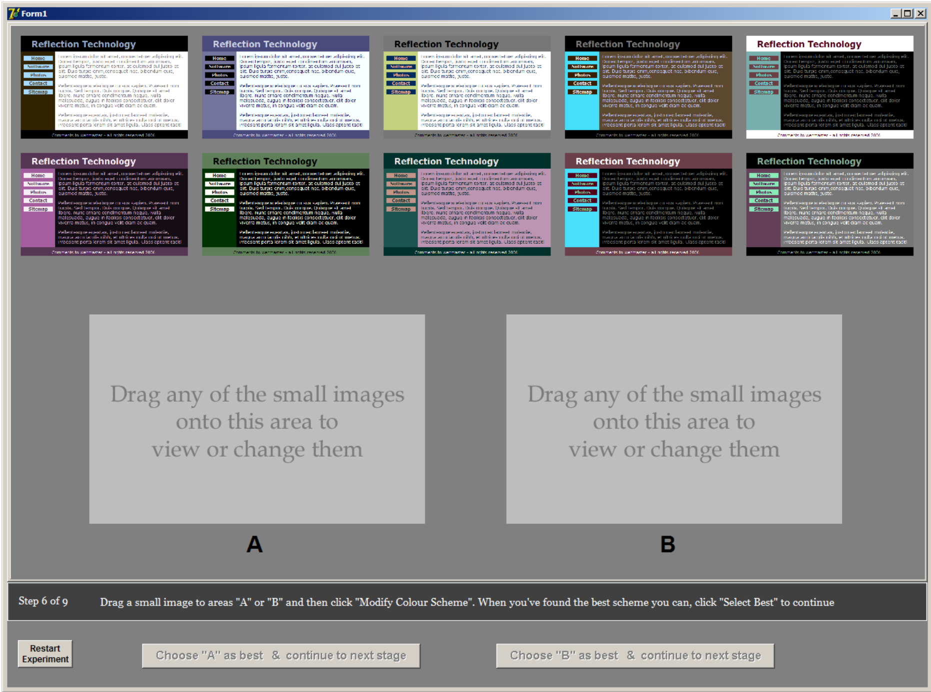


Figure 6.7: Harmoniser-based colouring: the initial view displays a set of randomly selected Colour Harmoniser-created schemes as "starter" schemes for the user to adjust. The user can drag any thumbnailed scheme to the lower A or B areas to have the scheme enlarged, so subtle colour differences can be seen more clearly.

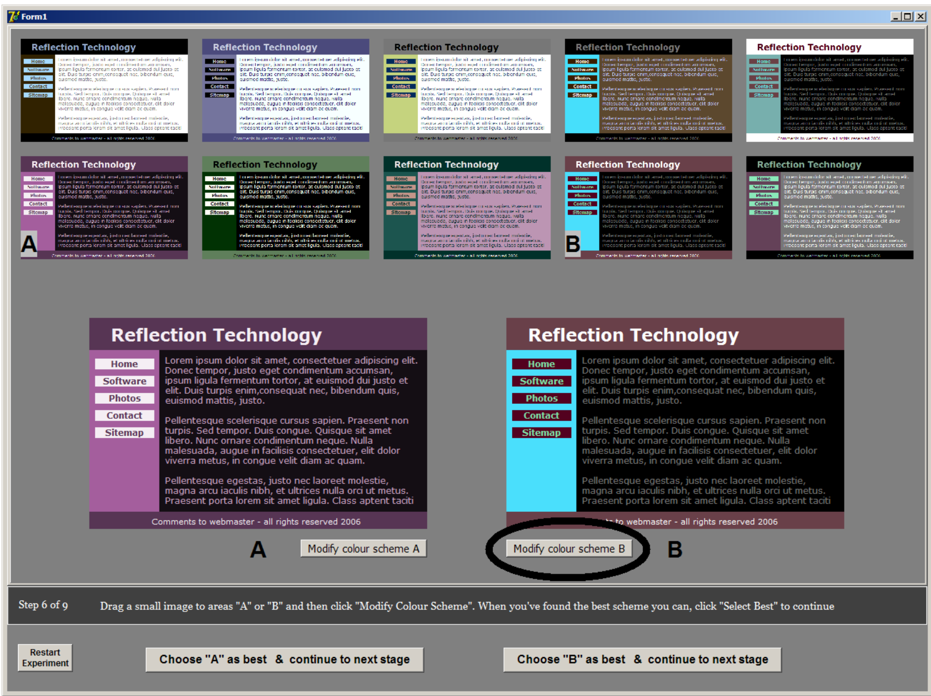


Figure 6.8: Harmoniser-based colouring: Dragging a thumbnail to the lower display areas will enable the "modify colour scheme" button, allowing the colour scheme to be modified. The A & B show the currently enlarged thumbnails.

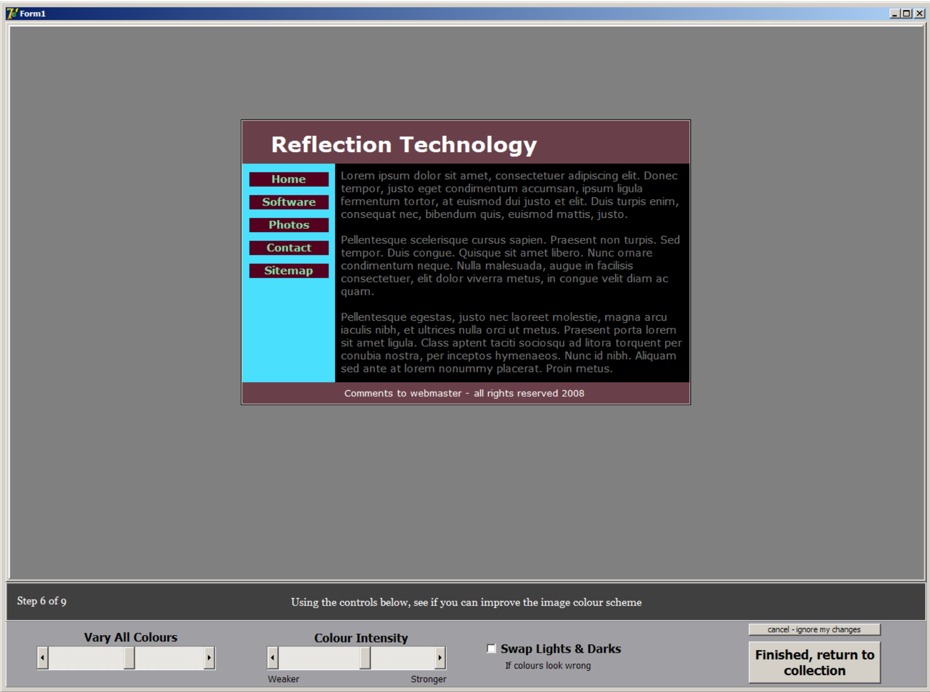


Figure 6.9: Harmoniser-based colouring: Using the holistic adjustment controls, users can alter a scheme by varying all the colours (wireframe rotation), altering the colour intensity (saturation) or swapping the light and dark colours. The non-modal interface immediately updates the colours as soon as any control is altered. The mid-grey background minimises simultaneous contrast effects.

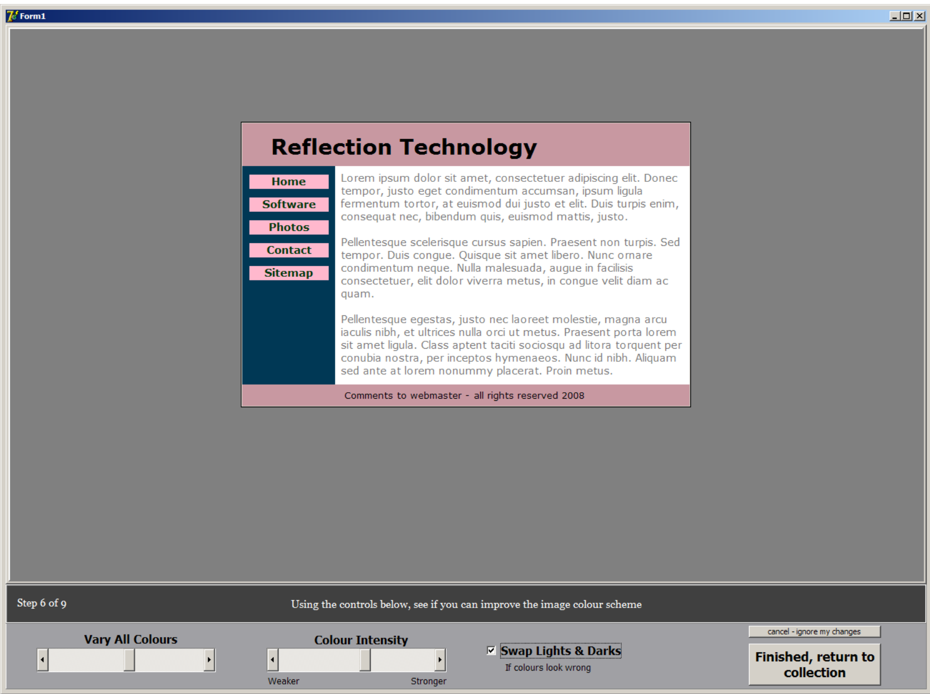


Figure 6.10: Harmoniser-based colouring: The holistic colour scheme adjustment interface showing the same colour scheme as the previous image, but with the lightness inverted.

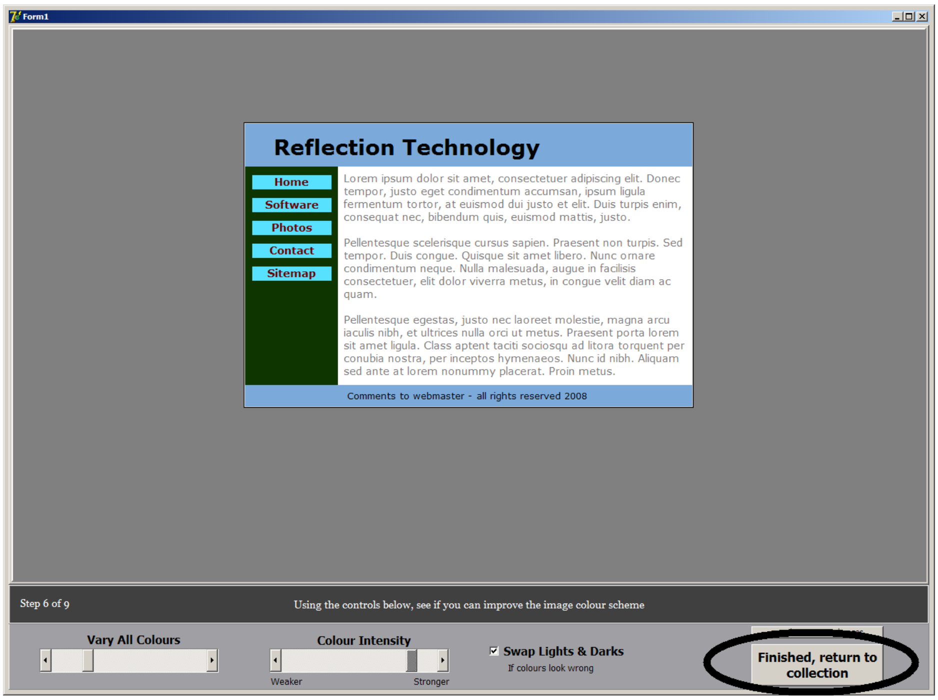


Figure 6.11: Harmoniser-based colouring: Showing the previous colour scheme after varying the colours and increasing the saturation. Clicking the indicated button to finish the colouring closes this window, and, after updating the colours of the enlarged image and its related thumbnail, redisplays the main interface.

schemes displayed in the enlarged area are preserved in the related thumbnail. Effectively, the thumbnails are the master-copies of the colour scheme, with the enlarged areas simply being a way of indicating a pair of selections. An illustration of the optimisation of a Harmoniser-based colour scheme (the "All-colours-together" experimental interface) is shown in figures 6.6 to 6.12.

The post-colouring and comparative questionnaires

At the completion of colouring a scheme (for both methods), the participant completed a post-colouring survey. The surveys are shown together in figures 6.13 and 6.14, but during the trial they were administered after the respective activities: i.e. the manual-colouring, then the manual post-assessment survey (fig. 6.13); the Harmoniser-based colouring, then the Harmoniser-based post-assessment survey (fig. 6.14); followed by the side-by-side comparison (fig. 6.15), and concluding with the questionnaire comparing both methods (fig. 6.16).

The side-by-side display of the two schemes and the instruction to "click on the image with the more professional colour scheme" (fig. 6.15) enabled the participant to make a direct comparison of the quality of the their schemes. This gives a first-order (although possibly biased) indication of the quality of the colour schemes created using the two methods.

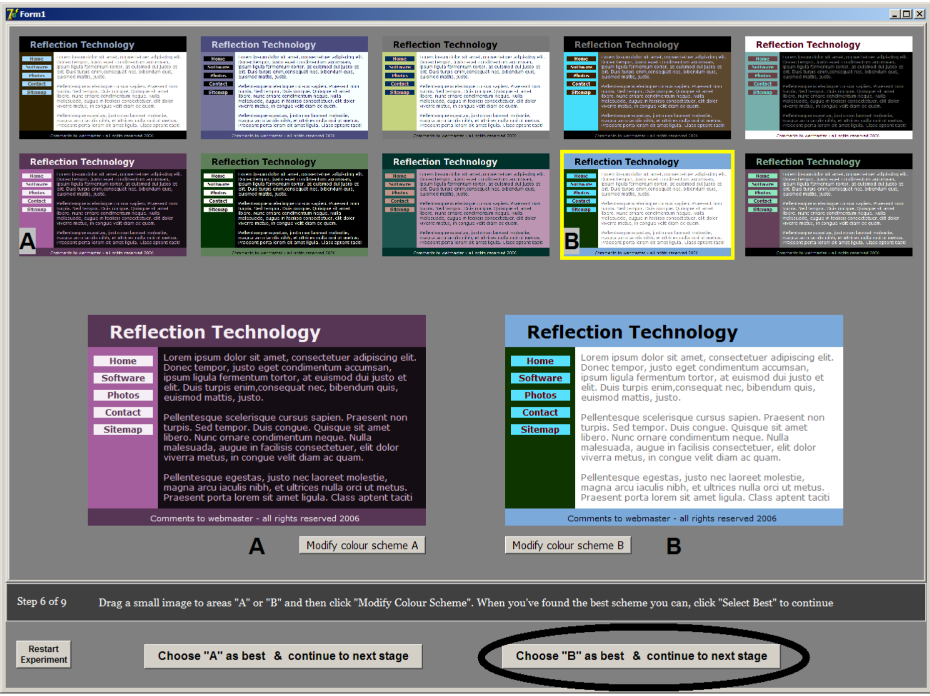


Figure 6.12: Harmoniser-based colouring: On return from the adjustment interface, the selected scheme and its thumbnail are updated to the colours chosen in the colour-change interface. Different thumbnails can be dragged to the lower A/B positions to explore other colouring and to see subtle colour differences more clearly. Clicking one of the "Choose A/B as best" buttons ends the colouring and displays the post-colouring questionnaire.

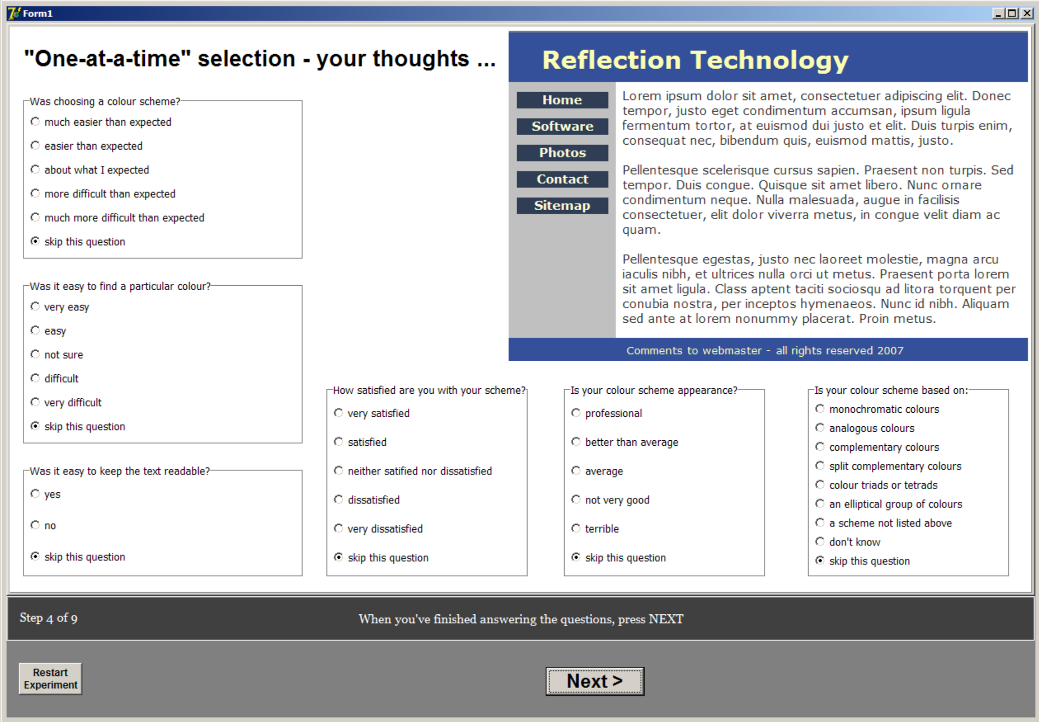


Figure 6.13: Manual colouring trial: having created their colour scheme, the user is shown the interface and asked to evaluate the process and the result.

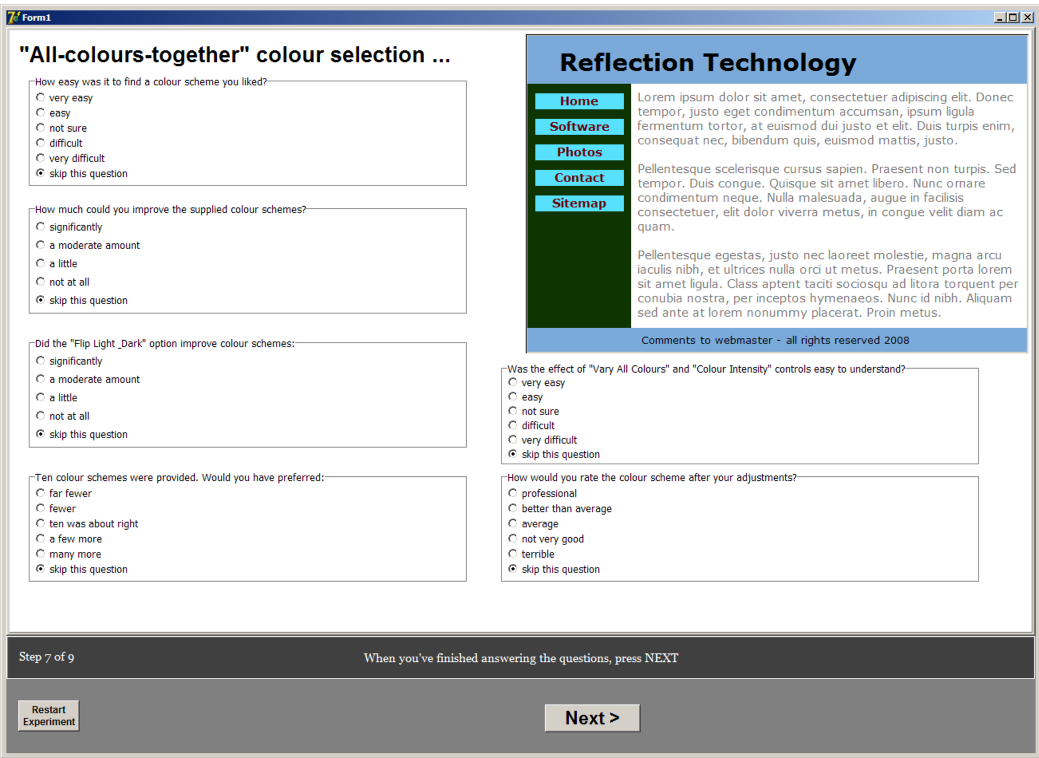


Figure 6.14: Harmoniser-based colouring: having adjusted a Harmoniser-based scheme, the user asked for their opinion on the process and the result.

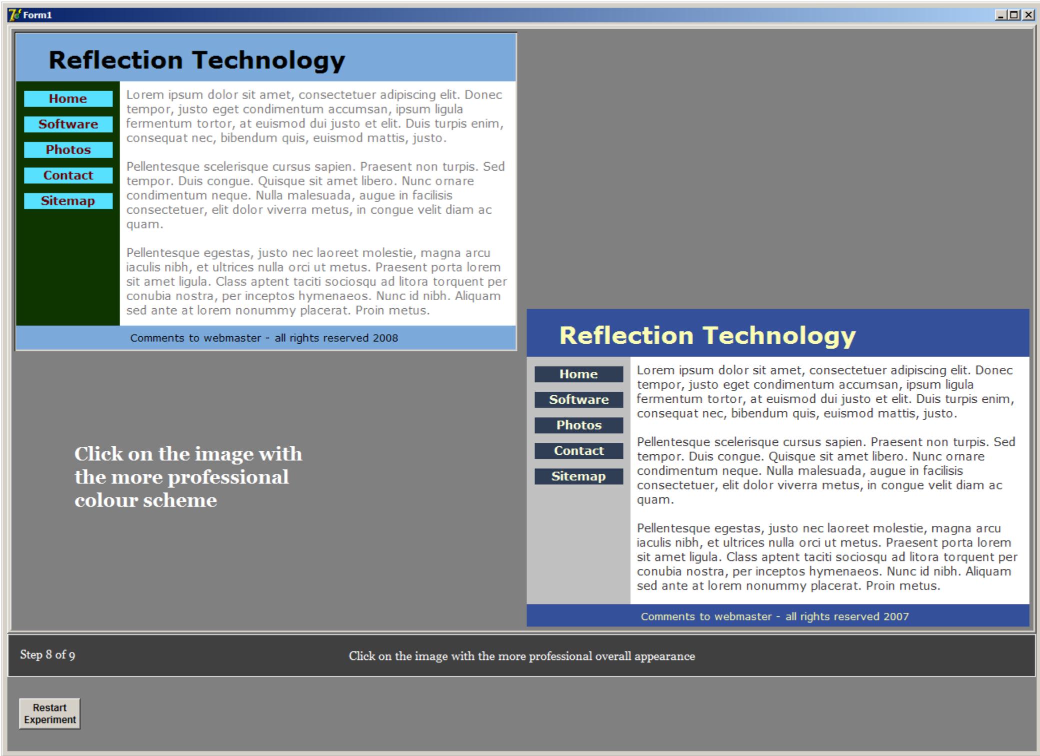


Figure 6.15: Having created two schemes, both are displayed and the user is asked to select the one with the more professional colour scheme.

6.1.5 The results of the comparative colour scheme creation experiment

The participants in this trial were students from a local girls' high school, staff and students from Massey University, family, and friends. One participant ignored the manual colouring section of the experiment, so their data was excluded from the analysis, leaving 73 participants (39 female, 34 male). The results are presented in two sections: the first contains results pertaining to the perceived quality of the resultant colour schemes, while the second contains an analysis of data relating to usability.

Rating the appearance of the resultant colour schemes

In the post-colouring questionnaires, the participants were asked to rate their manually-created scheme and their optimised Harmoniser-based scheme on five-point Likert scales, from "terrible" to "professional". Visually, the results (figure 6.17) show a slight bias in favour of the optimised Harmoniser-based scheme, but the difference is not statistically significant (t-test: t=-1.006, df=140.6, p=0.32). This is encouraging: the participants considered that the algorithmically-created schemes with their unchangeable relationships between the item colours were, after optimisation, comparable to those that they had created themselves. There were no effects relating to gender, age or the participant's self-perception of colour sense or of being an artist.

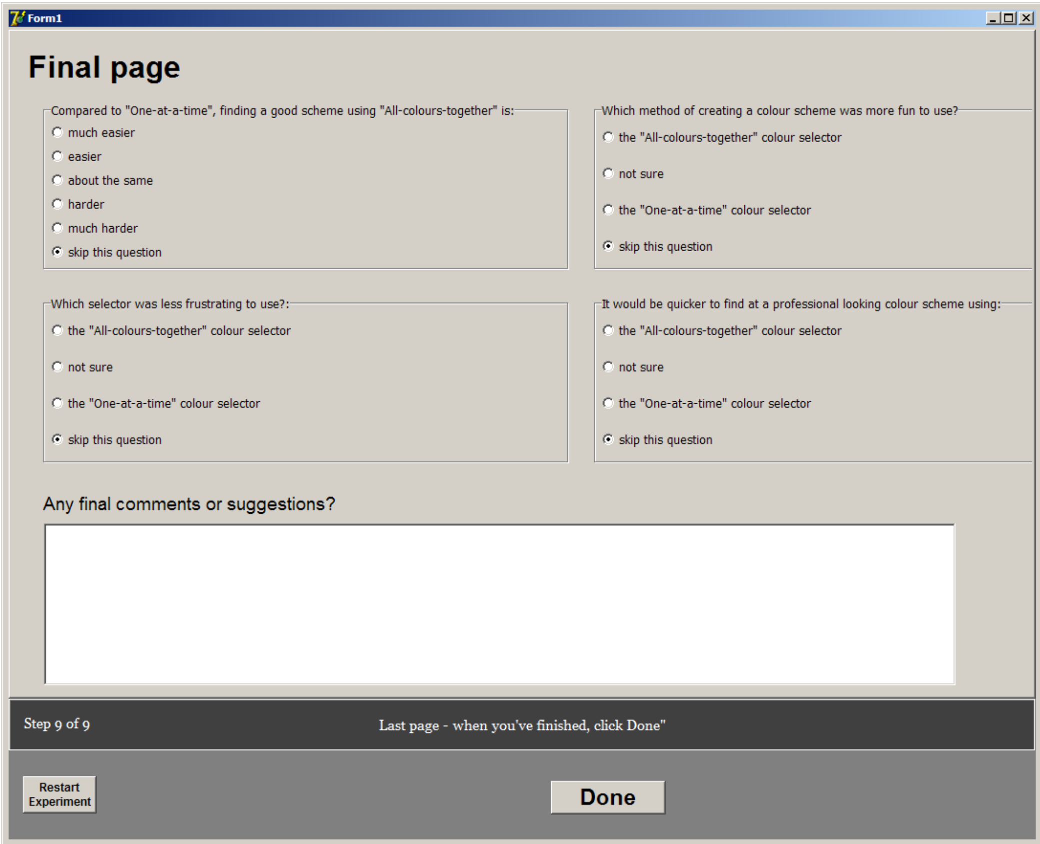


Figure 6.16: In the concluding questionnaire, the user is asked to compare both methods of colour scheme creation.

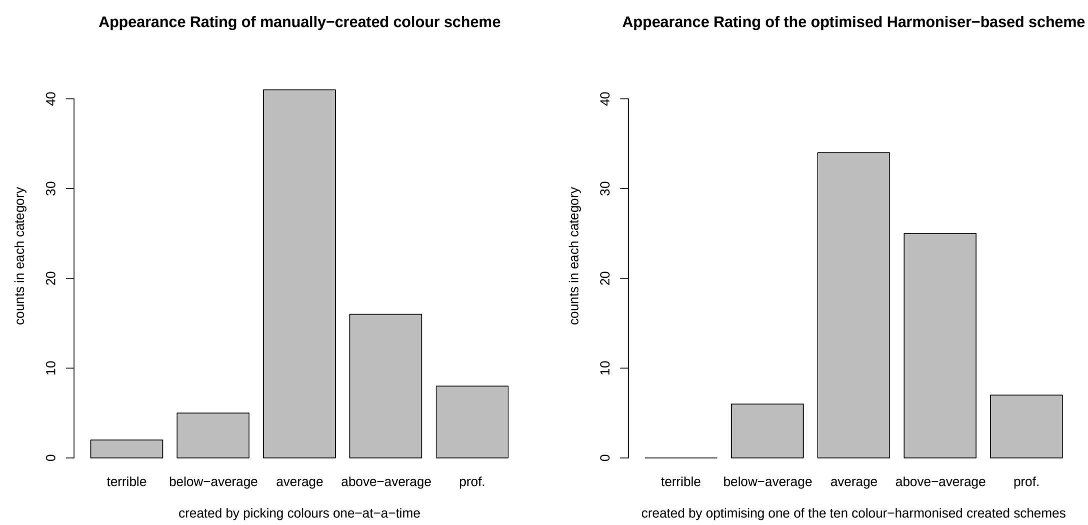


Figure 6.17: Rating the final colour schemes, one prepared by choosing individual colours for interface elements, the other by holistically adjusting a Harmoniser-based scheme.

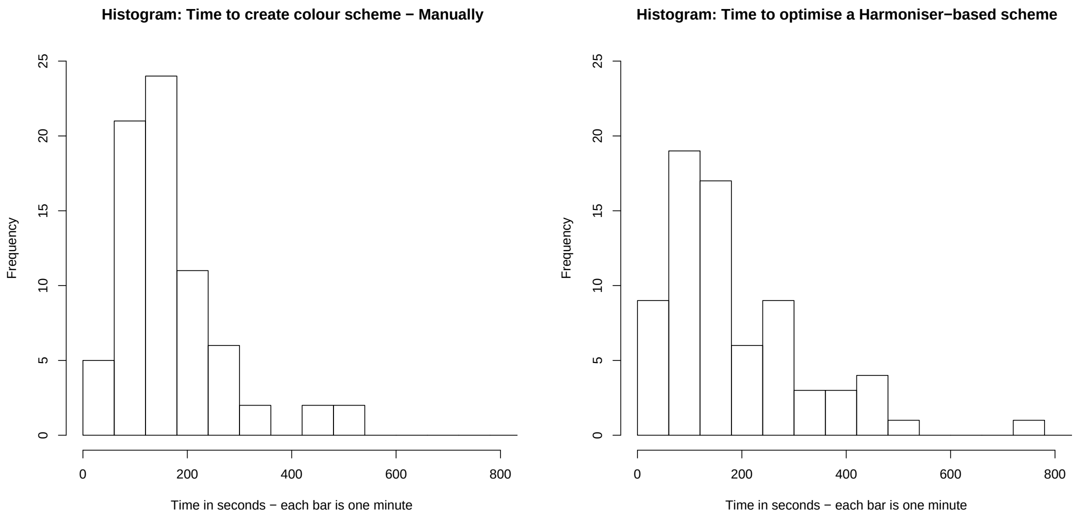


Figure 6.18: A histogram of the completion times for creating a colour scheme manually (left) and optimising a Harmoniser-based scheme (right). Each bar is one minute. Most participants took five minutes or less whichever method was used.

Choosing the more professional colour scheme

Having completed the creation of a scheme using both methods, the participants were shown both colour schemes, side-by-side (fig. 6.15), and asked to select the one with the more professional[4] scheme: 43 of 73 favoured the optimised Harmoniser-based scheme. This difference between the chosen options is not statistically significant (binomial test 43/73, p = 0.160). This is consistent with there being little difference between the assessments at the end of each individual colouring. At the 5% level, there was no gender effect in the preference of the colour scheme resulting from one method over the other (χ^2 test).

Thus far, the results indicate that creators of colour schemes consider that the quality of colour schemes they produced by adjusting raw schemes created by the Colour Harmoniser is not statistically different from the quality of schemes they created by colouring interface elements individually.

Comparing the time to completion

Using either method, most users took less than five minutes to choose a colour scheme. There were some outliers[5] who took much longer (fig. 6.18), but there was no significant difference between the completion times of the two methods (t-test: t=-0.007, df=115.5, p=0.995). However, as shown next, the users' expectation was that by using the Colour Harmoniser method, a professional scheme would be arrived at more quickly.

[4] it was felt that "professionalism" encompassed many of the elements necessary for a colour scheme to be usable and visually appealing, and was a term that would not require explanation.

[5] one of whom commented that they really enjoyed playing with the selectors.

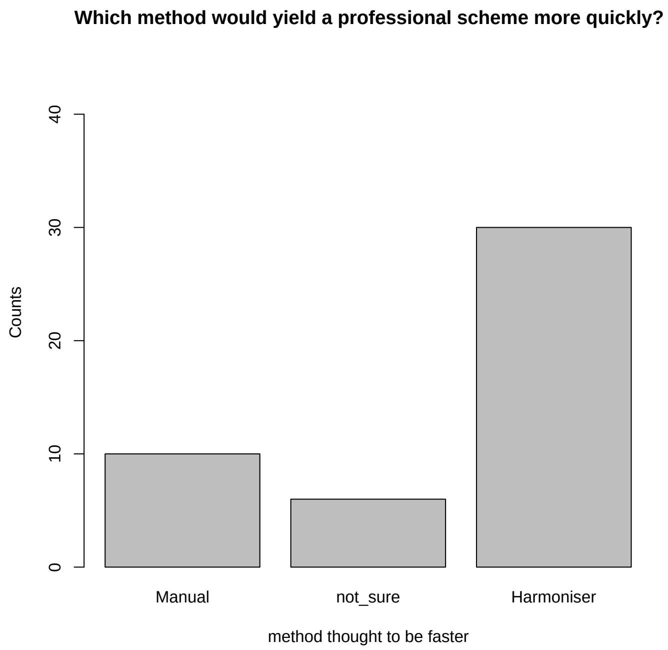


Figure 6.19: Responses to "would it be quicker to find a professional looking scheme using the manual or "all-colour-at-once" (Harmoniser-based) selector?"

The speed of finding a professional-looking scheme

The responses to the question *"It would be quicker to find a professional looking scheme using" (manual/not sure/Harmoniser)* was unambiguous. The results (fig. 6.19) are statistically significant, with 30 voting for the Harmoniser-based method, and 10 for the manual method (binomial test 30/40, p= 0.0022).

This is interesting. The participants, when creating a scheme themselves were undecided as to which method gave more professional results, but thought that, generally, the Colour Harmoniser-based approach would yield a professional looking scheme more quickly. This could be interpreted as meaning that, while they thought they could possibly do as well as the Harmoniser-based method, it would take longer.

Between-group consistency: This question about speed of finding a professional scheme was a late addition, being added to the concluding questionnaire after some trials (27 of 73) had been completed. Therefore, only the last 46 participants were presented with that question.

As the two sets of data were collected separately, as a consistency check, the two groups (first 27 vs. last 46) were compared to see if there was any significant difference in either their ratings of the appearance of the completed colour schemes (manual and Harmoniser-based), or in the proportion voting for the Harmoniser-based scheme in the side-by-side "which is more professional" test. Comparing the participant's rating of their schemes (manual and Harmoniser-based) in the post-creation questionnaire, there was no statistically significant difference between the two groups (χ^2 = 3.065, df = 3,

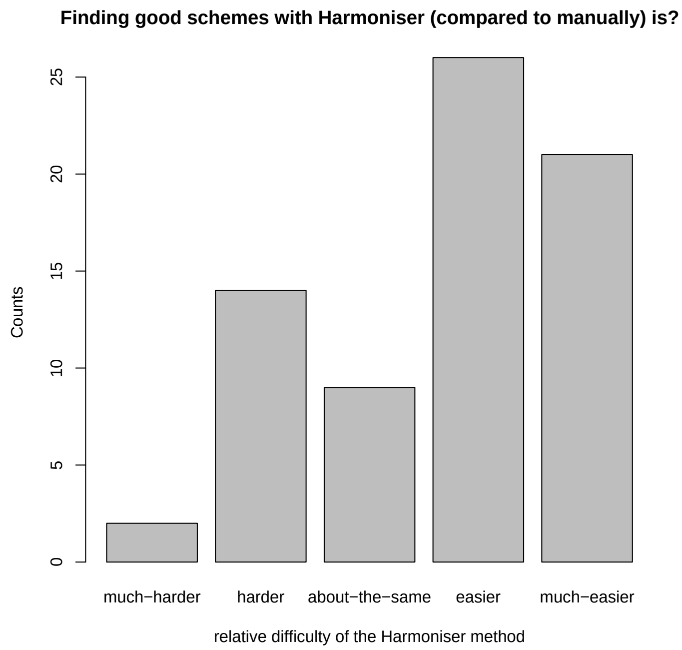


Figure 6.20: The responses to a question asking whether it would be easier to find good colours schemes using the Harmoniser method. It is clear that the majority of respondents think that finding good schemes would be easier using the Harmoniser method than by choosing colours individually.

p-value = 0.38). In the results of the side-by-side comparison of the preferred scheme, the differences in the proportions of those preferring the Harmoniser-based scheme in the first group (14/27) was not statistically different from the proportions of those in the second group (29/46), as given by the proportion test: p-value = 0.49.

Evaluations relating to the manual colouring method

From prior experience, it was expected that the participants would find two aspects of the task difficult: choosing appropriate colours while keeping the text readable, and finding particular colours. The first expectation was borne out, 62 of 72 users indicated that keeping the text readable did complicate colour selection (binomial test, $p < 0.001$).
The other expectation, that finding colours would be difficult, was not supported. Of the 72 responses, 37 found it easy or very easy to choose colours, whereas only 21 found it difficult, and only 2 very difficult. Twelve were undecided. The mean score was 3.4, above the "not sure" value of 3 and towards the "easy" end of the scale. The difference above 3 ("not-sure") is statistically significantly (t-test: t=2.74, df=71, p=0.0078).

Evaluations relating to the optimisation of Harmoniser-based schemes

The responses to questions about the usability of the method for choosing and adjusting colour schemes based on a Harmoniser-created scheme were generally positive.

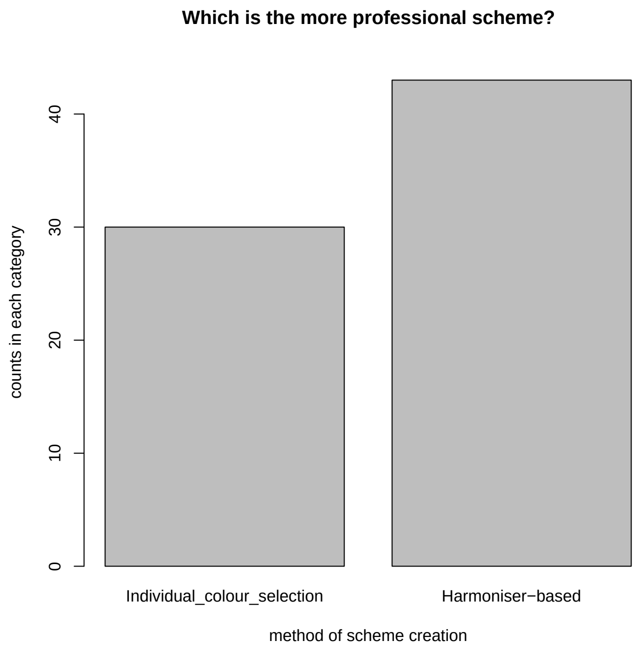


Figure 6.21: Which colour scheme is more professional? – the responses after the participants had created a colour scheme using each of the manual and Harmoniser-based methods. The difference between the assessments of the resulting schemes is not statistically significant, in contrast to the results shown in figure 6.19

Participants found that it was

- reasonably easy to find a scheme they liked (one-sided t-test: t=1.65, df=70, p=0.052),
- easy to understand the effect of the "vary all colours" and "colour intensity" controls (one-sided t-test: t=13.29, df=71, $p < 0.001$), as seen in figure 6.22,
- possible to improve the schemes using the holistic controls, as can be seen from figure 6.23;
- beneficial to use the "flip-light-dark" control, as can be seen from figure 6.24.

Comparing the usability of manual and Harmoniser-based methods

Three more questions relating to usability were included at the end of the trial. Having indicated that they expected the Harmoniser-based method would enable professional looking schemes to be found more quickly, the participants also indicated that

- finding a good colour scheme would be easier using the Harmoniser-based method, as shown in figure 6.20. The difference between mean of the results (3.7) and the neutral "about the same" value (3) is statistically significant (t-test: t=5.03, df=71, $p < 0.001$),

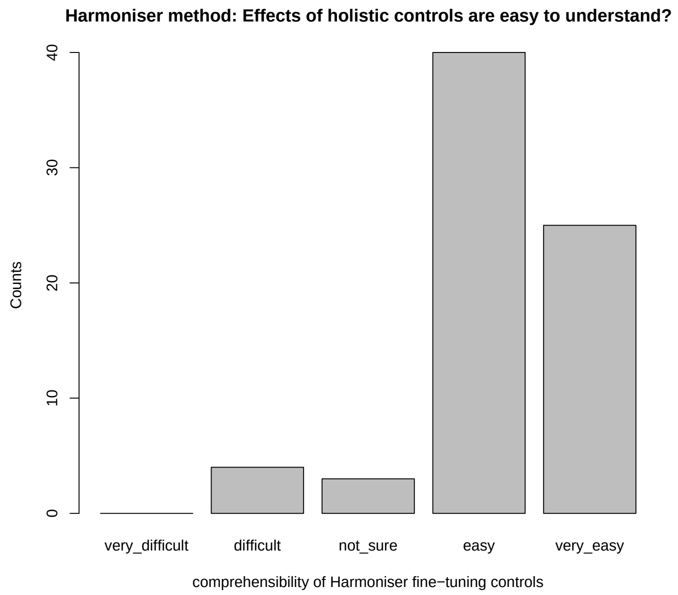


Figure 6.22: This chart shows the degree to which the users found the effect of the Harmoniser-based holistic colouring controls understandable.

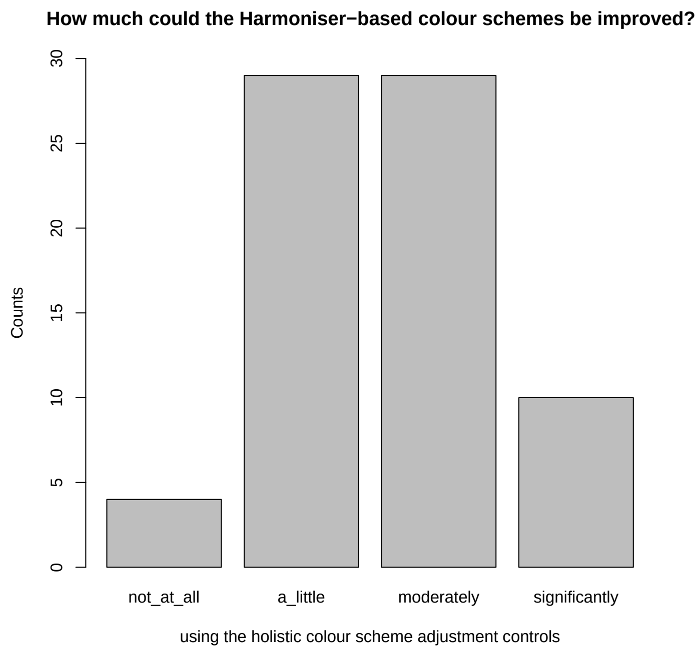


Figure 6.23: This chart shows the degree to which the users found that the Harmoniser-based schemes could be improved using the holistic controls.

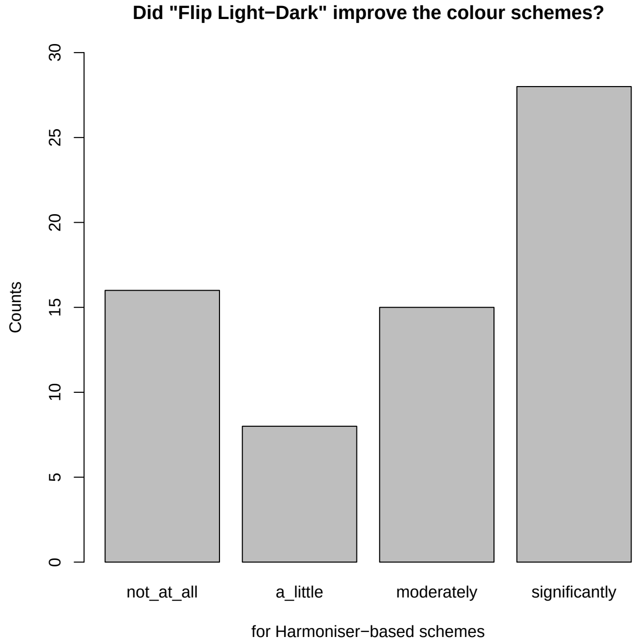


Figure 6.24: The flip-light-dark option inverts the lightness of colours in Harmoniser-based colour schemes. The responses indicate the degree to which the control were perceived to be beneficial.

- the Harmoniser-based method is less frustrating, although the result is of marginal significance (binomial test, 26 for harmoniser out of 41, $p = 0.059$),
- in terms of fun ("which was more fun to use?"), there was no significant difference: 25 for the Harmoniser out of 63 responses was not significant at the 5% level.

The comparative colour scheme creation experiment – a summary

When used for creation of user interface colour schemes with an "all-colours-at-once" holistic method, the Colour Harmoniser was found

- to be thought to make it easier to find good looking schemes,
- to be more likely to yield professional looking colour schemes,
- to have colouring controls that were easy to understand,
- to take comparable amounts of time to the manual method, and
- to produce results that were comparable in quality to colour schemes produced manually.

This is very positive. The colour schemes based on the predefined Colour Harmoniser designed colour relationships are comparable to schemes created by users to their own preference. The judgements also appear to be unaffected by the gender or colour sense of the evaluator, or previous colour training. While the time taken to create the schemes with the Colour Harmoniser was no shorter, the users expected the results to be more professional, even though this was not reflected in the results of their comparisons between the Harmoniser-based colour schemes and their own manually created schemes. The results of the next, larger-scale, experiment, in which participants scored colour schemes without being aware of their origins, suggest that there was an element of ownership-bias in these results.

6.2 Evaluating colour schemes created using the Colour Harmoniser

The primary purpose of the previous activity was to compare the ease of producing colour schemes using a Colour Harmoniser-based approach with a more conventional method of colour scheme creation, and to collect a set of adjusted Harmoniser-based schemes. The participants were also asked for their impressions of the quality of the colour schemes produced by the two approaches. However, because of personal investment in their own colour schemes, this group of participants would not produce a truly objective assessment of the quality of the colour schemes produced by the Colour Harmoniser. A more robust test was required, to enable an unbiased assessment of the quality of Harmoniser-based colour schemes relative to human-created schemes. In this test, it was therefore important that evaluators were unaware of the source of the schemes.

6.2.1 Aim of the experiment

The aim of this experiment is to measure the quality of colour schemes created by the Colour Harmoniser and several other methods in terms of several relevant quality criteria.

6.2.2 Experimental design

The experiment is a between-groups design, using images of colour schemes created using five different methods, including:

1. *schemes by artists:* 48 manually-created colour schemes by those who classified themselves as artists.
2. *schemes by non-artists:* 93 manually-created colour schemes by non-artists.
3. *Harmoniser-raw schemes:* 80 randomly-selected Harmoniser-created schemes. These were the unadjusted schemes created for use as the starter pool in the "Compare the Methods" experiment. For details, see section 6.1.3.
4. *Harmoniser-adjusted schemes:* 62 colour schemes resulting from the optimisation of the Harmoniser-based schemes by the participants in the previous "Compare the Methods" experiment.

5. *random:* 100 colour schemes with the same groupings[6] used when creating the Harmoniser-based schemes, but with the colour of the elements chosen randomly. The inclusion of the random schemes enables the conjecture that pleasing colour schemes are rare to be experimentally tested, and if this is so (as expected), the poor scoring randomly coloured schemes act as a reference for the bottom end of the quality scale.

Colour schemes randomly selected from the above groups were presented for evaluation individually. Each scheme was scored on criteria related to visual appearance and suitability for purpose. Demographic data for the participants, including details of previous colour design experience and factors relating to artistic self-perception, were also collected.

6.2.3 Experimental method

Participants were shown a set of interface colour schemes and asked to score each scheme on four criteria, each on a five-point Likert scale. The criteria were chosen to enable evaluations of:

- visual appeal (to give an assessment of the colour scheme independent of any particular use),
- professionalism (for on-line use, the professionalism of a colour scheme can be linked to perceived competence and trust, important issues for online business),
- how artistic (this is somewhat orthogonal to the use of a colour scheme for an interface, but was felt to be worth asking),
- the suitability of the colour scheme for use on a website.

Only the first criterion (visual appeal) had a pop-up hint. It displayed

> *This question enables you to indicate you really like the colour scheme, even if it doesn't appear professional.*

This is to allow for those colour schemes, for reasons that may be inexplicable, really appeal, even though they seem to break all the rules. The criteria were ordered as listed above, as the first two (visual appeal and professionalism) are related more to immediate impressions, and the "website suitability" question was placed last because it summarises the usability of the scheme.

Sources of schemes by artists and non-artists: The schemes by artists and non-artists were those created by participants using any colour they desired for any interface element without constraint – the manually created schemes. As part of the

[6] grouping the following so that each pair of items used the same colour for: the header and footer text; the header and footer background; the colour of the text on the buttons; and the buttons themselves.

demographic survey, each participant indicated whether they consider themselves an artist or not. The manually created schemes were separated into two groups depending on the answer to this question. The schemes in these groups were used as the manually created schemes by artists and non-artists.

The manually-created colour schemes included those created as part of the "Compare the Methods" experiment, and to increase the number and diversity, also schemes collected separately. The separate collection phase used a manual colouring interface that was virtually identical to that used in the "Compare the Methods" experiment[7].

The "Compare the Results" experiment requires users to assess Harmoniser-based colour schemes along with humanly created schemes. As schemes from CVD participants may use unusual colouring, those from known CVD participants were excluded from the artist/non-artist pool.

Each participant was shown fifteen colour schemes. The number of possible colour schemes is so large that the concept of a representative sample has little meaning. However, by including: colour schemes created by people from a diverse range of relevant demographic categories; colour schemes created by the Colour Harmoniser – both raw and adjusted – using different hues and colour scheme wireframes; and colour schemes coloured randomly, a wide variety of colourings is assured.

There are a different number of colour scheme images in each of the five classes, so to ensure an even sampling, one of the five classes of image was selected at random and a random image from that class selected, then one of the four remaining classes were selected and an image chosen. This was repeated until one image from all five classes had been evaluated. This sequence was repeated three times, giving an evaluation of fifteen randomly selected images with an even distribution between the classes, but without any fixed ordering.

There was one modification to this process. To enable the consistency of the individual evaluations to be determined, in 30% of trials, one colour scheme was shown twice: the colour scheme shown as the third was displayed for reassessment in position fourteen. The third image was used to reduce any "first evaluation" effects, and the fourteenth as it was sufficiently late in the trial that the use of a repeated scheme would not be too obvious, and the participant would be unlikely to remember their earlier assessment.

As in the previous trials, the computer-based trial commenced with a page of introduction that described the experiment, followed by a demographic survey. The fifteen colour schemes to be evaluated were displayed individually, and the trial concluded with a "thank-you page" (there was no need for a closing survey). The experimental interface used for the colour scheme evaluations is shown in figure 6.25.

The lighting could not be controlled, but in no case were the conditions atypical of those of a home or business environment.

[7] The only difference was the removal of a panel of buttons. The earlier interface had an additional method for changing the colours of interface elements, via buttons (one per element), that would invoke the colour selection dialog. However, as the buttons were rarely used, they were deleted from the manual interface used in the "Compare the Methods" experiment (fig. 6.2).

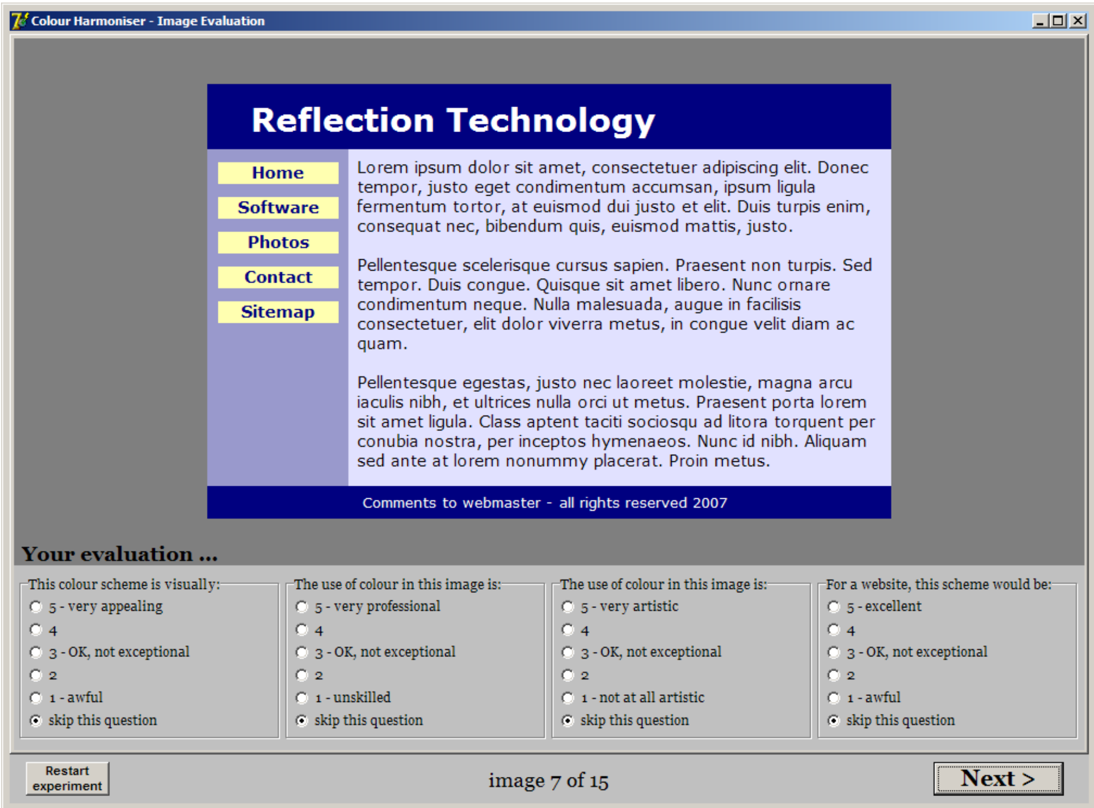


Figure 6.25: The colour scheme evaluation page used in the "Compare the Results" experiment.

6.2.4 Results

The participants in this trial had a variety of backgrounds: they included members of the public, design students, engineering and other non-artistic students in a variety of locations (passers-by/visitors to a museum, an art gallery, a design school foyer, University labs, a Linux user group meeting, businesses, and homes). In total, there were 127 participants, and the resultant data set had 1865 colour scheme evaluations. Not all participants completed all fifteen evaluations. Forty-six schemes were evaluated twice by the same participant for the consistency check. After removing these forty-six evaluations, the final data set had 1819 discrete evaluations, with around 360 for each method of creation. The demographics are shown in table 6.1. As the participants were free to skip questions, the totals do not always sum to 1819.

Gender	female 904 (50%), male 900, (50%)
Artist talent (self-assessed)	artists 603 (35%), non-artists 1127 (65%)
Age	5–14: 2%, 15–24: 28%, 25–34: 18%; 35–44: 14%, 45–54: 19%, 55–64: 13%, 65+: 6%
Colour training[8]	yes – 39%; no – 61%
Colour sense (self-assessed)	terrible (0, 0%), not very good (133, 7%), average (876, 49%), better than average (660, 37%), excellent (105, 6%)
Colour Vision	normal colour vision (1744, 97%), impaired vision (60, 3%)

Table 6.1: The demographic data from the colour scheme evaluations in the "Compare the Results" experiment.

The presentation of the results has been broken down into several stages. Firstly, an overview of the mean user assessments for the colour schemes created using each method is presented. Secondly, the final output of the Colour Harmoniser method – Harmoniser-based schemes that have been adjusted (personalised) – are compared to schemes created using the other methods. Then the raw (unadjusted) Harmoniser-based schemes are assessed to find out how these schemes, created without human intervention, compare with the schemes created by artists and non-artists. Checks for any demographic effects (e.g. gender and prior colour training) are included and any effects found are analysed in more detail.

To avoid confusion between the adjusted and unadjusted Colour Harmoniser-created schemes, the unadjusted schemes will often be referred to as the "raw" schemes (e.g. Harmoniser-raw, CHRaw). This wording will more clearly distinguish them from the Harmoniser-based schemes having the benefit of human adjustment (i.e. Harmoniser-

[8] Colour training: *e.g. have you been to art classes, read books on design or the use of colour...* to assess whether (formally or not), the participants had prior knowledge of colour design principles.

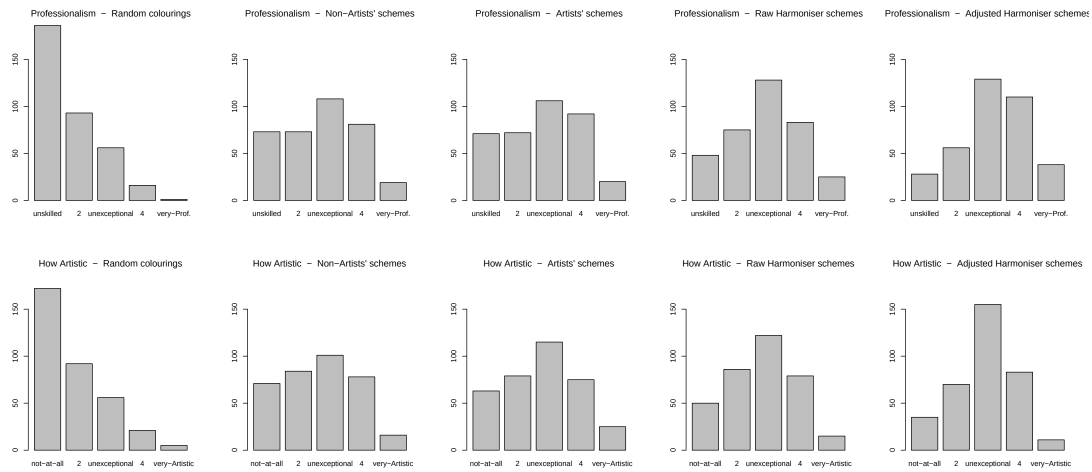


Figure 6.26: The participants' evaluation of how professional (upper) and how artistic (lower) the colour schemes were for each creation method. The categories within each plot indicate increasing quality from left-to-right: *unskilled* to *very professional* (top row), and *not-at-all* to *very artistic* (bottom row). From left to right, the columns of plots indicate schemes coloured: randomly, by a human non-artist, human artist, Harmoniser-raw, and Harmoniser-adjusted. There is a clear improvement in quality from left to right. An enlarged version of these images is included in Appendix B.1, p260.

adjusted, CHAdj schemes).

The rating of colour schemes grouped by creation method

Each image was evaluated on a five-point scale for each of the four criteria: *visual appeal*, *professionalism*, *how artistic* and *suitability for website use*. The Harmoniser-adjusted colour schemes achieved the highest mean scores in all four criteria. For schemes produced by the Harmoniser-adjusted method, the highest mean score was for the professionalism criterion and the lowest was for the artistic evaluation. As an illustration of the resultant data, the histogram of these two criteria are shown in outline in figure 6.26, and in more detail in the larger version with readable graph legends in Appendix B.1, p260. The mean scores for the colour schemes in each class is shown in table 6.2.

The significance of the differences between the methods

As can be seen from Table 6.2, there is a clear ordering to the means. The schemes from the Colour Harmoniser, after adjustment, score more highly than schemes created using any of the other methods, including those of both artists and non-artists. From the markedly low scores of the randomly colouring schemes, it is clear that schemes created using this method are definitely towards the unacceptable (1 corresponds to "awful" or "without merit") end of the scale.

It is easy for colour schemes to look really bad, but much harder for them to look

	Evaluation scores			
Source of Scheme	Visual appeal	Professionalism	How artistic	Web suitability
Harmoniser Adjusted	μ=3.15 σ=1.01	μ=3.20 σ=1.07	μ=2.90 σ=0.97	μ=3.12 σ=1.10
Harmoniser Raw	μ=2.91 σ=1.10	μ=2.89 σ=1.12	μ=2.78 σ=1.08	μ=2.80 σ=1.18
Human Artists	μ=2.83 σ=1.24	μ=2.77 σ=1.19	μ=2.78 σ=1.17	μ=2.75 σ=1.24
Human Non-artists	μ=2.80 σ=1.23	μ=2.72 σ=1.18	μ=2.67 σ=1.16	μ=2.64 σ=1.25
Random Colouring	μ=1.74 σ=0.95	μ=1.73 σ=0.91	μ=1.83 σ=1.00	μ=1.56 σ=0.84

Table 6.2: The evaluation scores of the different methods of creating colour schemes. The participants rated the schemes using a five point scale. In the table, the scores range from 1 (worst) to 5 (best). Notable are the surprising placement of Harmoniser-based schemes above the humanly-created schemes, including those produced by artists, the improvement in scores caused by adjusting Harmoniser-based schemes, the consistency of ordering of scores for the image creation methods, and the poor scores of randomly-coloured schemes.

extremely good, and this is reflected in the mean scores shown in Table 6.2. The differences between the means scores are not large, but there is a definite trend in the orderings. The Harmoniser-adjusted schemes consistently have the highest mean scores, the next lower are the Harmoniser-raw schemes, then the manually created schemes (both artist and non-artistic), with the lowest mean scores being awarded to the randomly coloured schemes. The presence of an effect due to the method of scheme creation is reinforced by the consistency of the changes in proportions in the quality assessments of the schemes created using the different methods in figure 6.26. Each row of the five plots show the quality assessment for the five different methods of creating a colour scheme. Within each plot the five bars indicate the quality assessments, with the quality increasing from left to right.
For the five columns within figure 6.26:

- *column 1 – the randomly created schemes:* as can be seen, the assessments are strongly weighted towards the low quality end (the left).
- *column 2 – schemes by non-artists:* the assessments are fairly even across the five quality assessment, except for the scarcity of schemes in “very professional” category (top row) or in the “very artistic” schemes (bottom row).
- *column 3 – schemes by self-professed artists:* the balance is slightly more in favour of the higher quality schemes.
- *column 4 – Harmoniser-raw schemes:* the professionalism (top row) scores are better with fewer lower scoring schemes. However, the “how artistic” assessments are much the same as the column 3 (schemes by non-artists).
- *column 5 – Harmoniser schemes after human adjustment:* the assessments are clearly weighted towards the higher quality (right-most) end, especially for the professionalism criterion (top row).

Therefore, although the differences in the mean scores are not large, there is a clear trend of the colour schemes being of higher perceived quality when moving from randomly created schemes, to those created by non-artists and artists, Harmoniser-raw, with the highest overall ratings being given to Harmoniser-based schemes after adjustment.

The differences between mean scores of schemes created using the different methods was tested for statistical significance using t-tests. The results are shown in table 6.3[9] (the complete table is shown in appendix A.1). When statistical probability values are shown, an indicator of significance is included to give a visual indication of the level: *** for $p<0.001$; ** for $p<0.01$; * for $p<0.05$; and a dot for $p<0.1$.

Visual Appeal – t-test significance

	CH-Adjusted	CH-Raw	Artist	NonArtist	Random
CH-Adjusted	–	p=0.003 **	p<0.001 ***	p<0.001 ***	p<0.001 ***
CH-Raw	p=0.003 **	–	p=0.316	p=0.194	p<0.001 ***
Artist	p<0.001 ***	p=0.316	–	p=0.778	p<0.001 ***
NonArtist	p<0.001 ***	p=0.194	p=0.778	–	p<0.001 ***
Random	p<0.001 ***	p<0.001 ***	p<0.001 ***	p<0.001 ***	–

Professionalism – t-test significance

	CH-Adjusted	CH-Raw	Artist	NonArtist	Random
CH-Adjusted		p<0.001 ***	p<0.001 ***	p<0.001 ***	p<0.001 ***
CH-Raw	p<0.001 ***		p=0.159	p=0.041 *	p<0.001 ***
Artist	p<0.001 ***	p=0.159		p=0.533	p<0.001 ***
NonArtist	p<0.001 ***	p=0.041 *	p=0.533		p<0.001 ***
Random	p<0.001 ***	p<0.001 ***	p<0.001 ***	p<0.001 ***	

How Artistic – t-test significance

	CH-Adjusted	CH-Raw	Artist	NonArtist	Random
CH-Adjusted		p=0.121	p=0.121	p=0.004 **	p<0.001 ***
CH-Raw	p=0.121		p=0.950	p=0.184	p<0.001 ***
Artist	p=0.121	p=0.950		p=0.222	p<0.001 ***
NonArtist	p=0.004 **	p=0.184	p=0.222		p<0.001 ***
Random	p<0.001 ***	p<0.001 ***	p<0.001 ***	p<0.001 ***	

Web site Suitability – t-test significance

	CH-Adjusted	CH-Raw	Artist	NonArtist	Random
CH-Adjusted		p<0.001 ***	p<0.001 ***	p<0.001 ***	p<0.001 ***
CH-Raw	p<0.001 ***		p=0.607	p=0.076 .	p<0.001 ***
Artist	p<0.001 ***	p=0.607		p=0.214	p<0.001 ***
NonArtist	p<0.001 ***	p=0.076 .	p=0.214		p<0.001 ***
Random	p<0.001 ***	p<0.001 ***	p<0.001 ***	p<0.001 ***	

Table 6.3: The significance of the differences between the mean scores of colour schemes creating using different methods, estimated using a t-test (df∼700).
In all cases, the mean scores orderings are: *random* <*non-artist*<*artist*≤*CH-Raw*<*CH-Adjusted*. Therefore, if there is a significant difference between two methods, e.g. CH-Raw and CH-Adj, it means that CH-Raw<CH-Adj, or equivalently, CH-Adj schemes scored more highly and are therefore better.

[9] There are about 360 colour schemes in each class. The degree of freedom (df) for the t-tests reported here is around 700.

Assessing the merit of Harmoniser-created schemes, after adjustment

To enable an evaluation of the merit of the adjusted Colour Harmoniser-derived schemes, the data in Table 6.3 can be reformulated into a single table that shows the significance of the differences between the means of the adjusted Harmoniser-based colour schemes and those created using other methods (Table 6.4). As can be seen, for three of the four criteria, the Harmoniser-based schemes, with the benefit of human adjustment in the final colouring, rate more highly than all other methods of colour scheme creation, including, surprisingly, those of human artists.

Significance of differences from the Adjusted Harmoniser-created schemes

	Visual appeal	Professionalism	How artistic	Website suitability
Harmoniser-Raw	p=0.003 **	p<0.001 ***	p=0.121	p<0.001 ***
Artist	p<0.001 ***	p<0.001 ***	p=0.121	p<0.001 ***
Non-artist	p<0.001 ***	p<0.001 ***	p=0.004 **	p<0.001 ***
Random	p<0.001 ***	p<0.001 ***	p<0.001 ***	p<0.001 ***

Table 6.4: The significance of the differences between the means of the Harmoniser-created schemes *after human adjustment* and all other methods, for the four evaluation criteria. Orderings of the means: *random<non-artist<artist≤Harmoniser-Raw<Harmoniser-Adjusted*.

This superiority does not extend to the "how artistic" criterion, where there is no statistical difference between the Harmoniser-adjusted schemes, Harmoniser-raw schemes and human-artists. The adjusted Harmoniser-based schemes are, however, rated as being better than those of non-artists.

It may be that the colour scheme of the example website is too simple for the distinction artistic/non-artistic to be meaningful, no matter what colour scheme is used, and that such differences would only become apparent when more detailed elements such as textures and images are included. While not part of this research project, methods for their inclusion are discussed in section 7.5. Nevertheless, the Colour Harmoniser-based colour schemes are clearly visually appealing, do appear professional, and would be suitable for use online.

The scores of the four evaluation criteria were checked for any sensitivity to the evaluator's gender, age, previous colour training, colour sense, self-perception as an artist, and any known colour vision deficiency (Table 6.5). The only factor significant

	Visual appeal	Professionalism	How artistic	Website suitability
Gender	p=0.028 *	p=0.013 *	p=0.008 **	p=0.21
Colour Sense	no effects at 5% level			
Colour Training	no effects at 5% level			
Evaluator is Artist	no effects at 5% level			
Age	no effects at 5% level			
CVD	no effects at 5% level			

Table 6.5: Effects of demographics and past experience on the evaluations of the *Adjusted Harmoniser-based schemes*. No effects were apparent except for gender: females gave higher scores than males, across all four evaluation criteria.

at the 5% level was gender: females gave significantly higher scores than males, for all criteria except website suitability, as can be seen from Table 6.6.

	Mean scores		t-test results	
	Male	Female	Significance	details
Visual Appeal	3.02	3.26	p=0.028 *	t = 2.21, df = 356
Professionalism	3.05	3.33	p=0.013 *	t = 2.49, df = 354
How Artistic	2.75	3.02	p=0.008 **	t = 2.67, df = 348
Website Suitability	3.03	3.17	p=0.209	t = 1.26, df = 353

Table 6.6: The details of the gender differences in scoring the Harmoniser-adjusted colour schemes. For all but website suitability, females gave significantly higher scores.

The merit of Harmoniser-based colour schemes, before adjustment

The high scoring colour schemes discussed in the last section were created by the Colour Harmoniser prototype using a model based on accepted colour harmony heuristics, and the schemes were adjusted (personalised) post-creation to best effect by a person. The schemes therefore combined artistic heuristics with human judgement. The raw Harmoniser-created schemes do not have the benefit of explicit human judgement. This section describes how schemes created by the Colour Harmoniser – the Harmoniser-raw schemes – compare to the quality of colour schemes created by people.

As can be seen from Table 6.7, the Harmoniser-raw schemes are sometimes ranked more highly than humanly-created colour schemes. Omitting the randomly coloured schemes (which are uniformly worse than all others), the ratings of the schemes can be grouped as shown in Table 6.8.

As with the adjusted Harmoniser-based schemes, the "How artistic" criterion is anomalous, but more surprising is that the colour schemes created by the Colour Harmoniser prototype, using whatever colours resulted from an arbitrary wireframe rotation, are rated (overall) as not statistically different from those schemes created by humans. On average, the professionalism of raw Harmoniser-based schemes was comparable to that of artists, and better than the professionalism of schemes by non-artists. For the other criteria (visual-appeal, how-artistic, website-suitability), the Harmoniser-raw schemes were comparable to those by humans, both artists and non-artists.

Significance of differences from Harmoniser-Raw schemes

	Visual appeal	Professionalism	How Artistic	Website Suitability
Harmoniser-Adj	p=0.003 **	p<0.001 ***	p=0.121	p<0.001 ***
Artist	p=0.316	p=0.159	p=0.950	p=0.607
Non-artist	p=0.194	p=0.041 *	p=0.184	p=0.076 .
Random	p<0.001 ***	p<0.001 ***	p<0.001 ***	p<0.001 ***

Table 6.7: The significance of the difference between the means between the Harmoniser-created schemes *before any human adjustment* and all other methods, for all four evaluation criteria (two-tailed t-test).
Orderings of the means: *random*<*non-artist*<*artist*≤*Harmoniser-Raw*<*Harmoniser-Adjusted*.

Evaluation criteria	**Colour schemes creation method – perceived ordering**
Visual Appeal	(Harmoniser-raw, artist & nonartist) < Harmoniser-adjusted
Professionalism	Non-artist < (artist & Harmoniser-raw) < Harmoniser-adjusted
How Artistic	No statistically significant differences between non-artist, artist, Harmoniser-raw & Harmoniser-adjusted
Website Suitability	(Harmoniser-raw, artist & nonartist) < Harmoniser-adjusted

Table 6.8: The relative placement and grouping of colour schemes created using the different methods. After human adjustment, the Colour Harmoniser-based schemes are rated more highly than all other methods except for the artistic criterion, where they are judged as comparable to those of artists and Harmoniser-raw schemes. The Harmoniser-raw schemes are comparable to those created by humans, both artists and non-artists.

	Visual Appeal	Professionalism	How Artistic	Website Suitability
Colour Sense	p=0.28	p=0.22	*p=0.036 **	p=0.34
Age	p=0.40	p=0.037 *	p= 0.044 *	p=0.28
Gender	no effects at 5% level			
Colour Training	no effects at 5% level			
Evaluator is Artist	no effects at 5% level			
CVD	no effects at 5% level			

Table 6.9: Harmoniser-raw schemes: the effects of evaluator demographics and past experience on the evaluations. Unlike the Harmoniser-adjusted schemes there is no gender effect, but there are effects relating to colour sense and age.

There were two demographic and design background effects (Table 6.9). Unlike the Harmoniser-adjusted schemes, for the Harmoniser-raw schemes, gender was not significant, but there is a mild age effect, and one relating to the evaluator's colour sense. Using Tukey's HSD (Honest Significant Difference) to tease apart the source of the effect: the only pair of age-ranges that had differences in mean scores significant at the 10% level were the 55-to-64 age group when compared to the 15-to-24 (p=0.034) and 35-to-44 (p=0.079) age groups. The mean score of the Harmoniser-raw colour schemes from the participants in the 55-to-64 age group was significantly lower than all other age groups. This demographic was well-represented; overall there were 363 observations of Harmoniser-raw schemes with 47 (13%) from the 55-to-64 age group. This group would have grown up in the 1960s and 70s, times of striking changes in popular art and colour due (in part) to the discovery of new dyes (Varley, 1980). This may account for the differences in what is seen as artistic, but further research would be required to support this supposition.

For the "How artistic" criterion only, there is an effect relating to a participant's colour sense (anova: p=0.036 *), but this effect is not present for the Harmoniser-adjusted schemes. Visually, the primary difference between the Harmoniser-raw and Harmoniser-adjusted schemes is the selection of more appealing hues, and which hues are light and dark. These effects are much more dramatic than the changes from adjusting the saturation. One theory to explain the effect relating to colour sense is that this group, having a more developed colour sense, could "see through" the possibly inappropriate hues to the "goodness" of the underlying chromatic relationships between the coloured objects and were therefore scoring the Harmoniser-raw schemes more highly than those with less attuned (average) colour sense. This was found to be completely wrong. On further examination of the data, it was found that those who asserted they had excellent colour sense allocated much lower scores to the raw Harmoniser-based schemes than those who rated their colour sense more conservatively, as shown in table 6.10.

	Colour sense – self-perception				
	terrible	not very good	average	better than average	excellent
"How Artistic" mean score for Harmoniser-raw	—	2.88	2.79	2.87	2.11

Table 6.10: Harmoniser-raw schemes were scored more poorly by evaluators with excellent colour sense. The difference is statistically significant.

One interpretation is that this group, while sensitive to what looks right, are also highly attuned to what does not look right, and are therefore downgrading schemes using inappropriate hues. This is supported by the lack of any colour-sense effect with the Harmoniser-adjusted schemes (Table 6.5) as these have been adjusted to use (from a human perspective) "appropriate" colours.

The schemes created using the Colour Harmoniser method, even when unadjusted, score quite well. The results indicate that Harmoniser-based schemes where no attention is paid to the final dominant colours (as is the case when the wireframe rotational angle is randomly chosen) are seen as professional, as suitable for online use, and as having visual appeal equal to that of colour schemes created by human creators, including those with design skills.

A comparison between the schemes of artists and the Colour Harmoniser

One of the initial goals of this project was to try to improve the quality of colour schemes available to those without good colour design skills. It was hoped that the schemes created automatically would be better than those created by those unskilled and might approach those with artistic ability. While it is gratifying that the Harmoniser-based schemes scored so well, it was surprising to find that the colour schemes by artists did not score more highly. The significance of the differences between colour schemes created by artists and the other methods is shown in Table 6.11. As can be seen, there is no statistical difference between the scores of colour schemes prepared by artists, by non-artists and by the Colour Harmoniser (before adjustment).

It was rather surprising that the schemes created by those who classified themselves as artists were not, by most evaluators, scored any differently from the schemes by non-artists or those from the Harmoniser-raw schemes.

To determine whether there were any demographic effects, (e.g. evaluators who were artists themselves might recognise the schemes of other artists), an analysis of variance

Significance of differences from the schemes created by artists

	Visual appeal	Professionalism	How artistic	Website suitability
Harmoniser-Adj	p<0.001 ***	p<0.001 ***	p=0.121	p<0.001 ***
Harmoniser-Raw	p=0.316	p=0.159	p=0.950	p=0.607
Non-artist	p=0.778	p=0.533	p=0.222	p=0.214
Random	p<0.001 ***	p<0.001 ***	p<0.001 ***	p<0.001 ***

Table 6.11: The significance of the difference between the means between the *schemes created by artists* and all other methods, for all four evaluation criteria (two-tailed t-test). Orderings of the means: *random<non-artist<artist≤Harmoniser-Raw<Harmoniser-Adjusted*.

was performed on the criteria scores against the demographic categories: gender, the evaluator being an artist, previous colour training, and the evaluator's (self-assessed) colour-sense. Only colour-sense was significant at the 5% level (see Table 6.12 for details). It is clear that there is no support for the idea that artists would recognise and score the schemes of other artists differently (anova: scores-of-schemes-by-artists ∼ evaluator-is-an-artist, p=0.58). However, the colour-sense effect requires further investigation.

	Visual appeal	Professionalism	How Artistic	Website Suitability
Colour Sense	p=0.0072 **	p=0.023 *	p=0.011 *	p=0.039 *
Gender	no effects at 5% level			
Colour Training	no effects at 5% level			
Evaluator is artist	no effects at 5% level			

Table 6.12: Schemes by artists: there are no statistically significant demographics effects, except for the colour-sense criterion. Those with (self-assessed) excellent colour sense appear to be able to differentiate something in the schemes created by artists (from anova: EvaluationCriterionScore ∼ ColourSense). However, this difference is not (as might be suspected) to rate the artist's schemes more highly. Surprisingly, as can be seen in Table 6.13, schemes by artists are awarded lower scores.

Schemes by artists are scored poorly by those with excellent colour sense

During the analysis of the evaluations of the Harmoniser-raw schemes, it was noted that those who asserted they had excellent colour sense allocated lower schemes for the "How Artistic" criterion than did participants with average colour sense. The same effect was noted for schemes created by artists. Schemes created by artists were (as with Harmoniser-raw schemes) awarded lower scores by those with excellent colour sense than by those with average colour sense. On a scale of 1 to 5, the mean score awarded by those with average colour sense was 2.94, far above the 2.09 awarded by those with excellent colour-sense. The difference is statistically significant (t-test: t=2.91, df=25.4, p=0.008 **), but the reason is difficult to discern.

The testing for an effect of average vs. excellent colour sense for the "How Artistic" criterion scores was extended to include the non-artist and Harmoniser-adjusted creation methods. The results for all four creation methods are shown in Table 6.13.

As can be seen, for the "How Artistic" criterion, only the Harmoniser-raw method (as previously discussed) and the schemes by artists have a colour-sense dependency. There are no colour-sense effects with the Harmoniser-adjusted or non-artists' schemes.

	Mean "how artistic" score colour sense		t-test results	
	average	excellent	Significance	details
Harmoniser-Adj	2.88	2.90	p=0.924	t=-0.096, df=22.1
Harmoniser-Raw	2.79	2.11	p=0.018 *	t= 2.56, df=22.1
Artists' schemes	2.94	2.09	p=0.008 **	t= 2.91, df=25.4
Non-artists'	2.76	2.52	p=0.370	t= 0.912, df=25.8

Table 6.13: Schemes by artists are scored more poorly by those with excellent colour sense than by those with average colour sense. Unlike lower scores for the Harmoniser-raw schemes, the lower scores for the schemes by artists are less easy to explain.

To determine whether the lower scores awarded by those with excellent colour sense was restricted to the "How Artistic" criteria, the analysis was extended to include the other three evaluation criteria. The results are shown in Table 6.14. All four criteria: visual-appeal, professionalism, how-artistic, and website-suitability show a colour sense effect, and, like the "How Artistic" scores for the Harmoniser-raw schemes, the scores awarded by those with excellent colour sense are uniformly lower than those with average colour sense.

	Mean evaluation score, by evaluator's colour sense		t-test results	
	average	excellent	significance	details
Visual Appeal	3.01	2.09	p=0.004 **	t=3.13, df=26
Professionalism	2.93	2.14	p=0.015 *	t=2.62, df=25.1
How Artistic	2.94	2.09	p=0.008 **	t=2.91, df=25.4
Website Suitability	2.91	2.14	p=0.020 *	t=2.47, df=25.5

Table 6.14: The schemes by artists are given significantly lower scores by those with excellent colour sense than by those with average colour sense. This difference only exists for the artists schemes and for the "How Artistic" criterion of the Harmoniser-raw schemes.

While those with excellent colour sense may be downgrading "wrongly coloured" Harmoniser-raw schemes, no such explanation can apply here. The schemes that were prepared by people who have classified themselves as artists have been singled out and scored more poorly by those who indicated they have excellent colour sense. This effect cannot be caused by the Colour Harmoniser, as the manually created schemes were created by humans selecting their preferred colours. Apart from using the same website interface, the schemes are unrelated to the Colour Harmoniser. During the experimental trial, the schemes from artists and non-artists were scored individually, without the evaluator being aware of their origin. No reason is evident for the disparity in scores, and as this relates to human-created schemes being evaluated by other humans, it is interesting, but does not affect the results relating to the Harmoniser-based schemes.

Therefore, the reason for the lower scoring of the schemes by artists given by those excellent colour sense is left as a topic for future research.

Randomly coloured schemes are significantly worse than designed schemes

As can be seen from the mean scores in Table 6.2, p226, the randomly coloured schemes are scored by evaluators well below schemes created using any other method (including those by people who deem themselves non-artists), and the difference is statistically significant ($p < 0.001$ – from Table 6.3, p227). The poor scores for randomly coloured schemes – the mean score across the four evaluation criteria was 1.71 on a scale of 1 to 5 – is experimental validation of the rarity of harmonious and usable colour schemes.

The consistency of image evaluation

There were a small number of repeated observations during the image evaluations. Forty-six images were evaluated twice within a single trial. This was accomplished by presenting the colour scheme displayed as the third for reevaluation at position fourteen (out of fifteen). The colour scheme creation method for repeated schemes was evenly distributed over the five possible creation methods. There was a slight increase (μ=0.25 over the four criteria) in rating for the second occurrence of a colour scheme. This drops when the random schemes are omitted to μ=0.190, which is a slight but statistically significant increase (t=2.21, df=141, p=0.029). The slight increase in the second scorings could be due to the participants realigning their standards as more schemes are evaluated. Further research would be required to check this hypothesis.

The time required to judge an image

The median time taken to evaluate an image was 18 seconds, the first and third quartiles being 11 and 28 seconds respectively. Therefore, the average time to assess the score for each criterion was about four seconds. This suggests that the ratings were based more on initial impressions than deep analysis.

6.3 Discussion

At the start of the project, it was hoped that the post-adjustment Harmoniser-based schemes might rate more highly than schemes produced by non-artists, but it was expected that they would rate significantly lower than schemes produced by artists. However, this expectation turns out to have been too modest; for most criteria, the Harmoniser-based schemes score more highly than those created by people, including people who class themselves as artists. This was most unexpected, as was the high scoring of the raw Harmoniser-based schemes.

The results demonstrate that the proposed method of creating user interface colour schemes – combining colour harmony heuristics with constraints related to usability and user interface semantics – is both viable and robust. The method does not require specialist knowledge in order to create schemes, nor to personalise them once created.

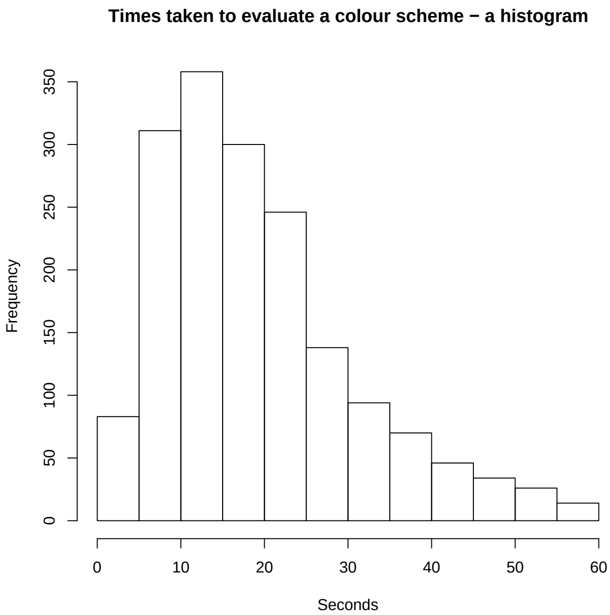


Figure 6.27: The times required to rate a colour scheme for all four evaluation criteria, with the data trimmed to exclude values greater than sixty seconds. This indicates that most users took only a few seconds to score each criterion.

The colour schemes produced by applying the extended model of colour harmony suggested in Chapter 3 result in schemes that are seen as visually appealing, professional and suitable for online use across a wide cross-section of users. Heartening though those results are, it is not claimed that those rules are optimal; they could be improved or augmented, or even replaced. For example, colour schemes for websites for young children would require different aesthetic factors in the colour balance model (see sec. 2.9.6), as the preferences of young children differ from teenagers and adults. There may also be circumstances in which colour disharmony is required; just as dissonant mood music is used in films to suggest tension and anxiety, so disharmonious colour schemes could be used to evoke feelings of unease or even repugnance where necessary. Clearly, a different set of rules would need to be used (and designed) to generate such different effects. The optimiser is indifferent to the rules that are plugged into it; if a set of rules favouring saturated colours, or disgusting colour combinations is plugged into it, then the optimiser will happily generate such schemes, while still ensuring their usability, and allowing them to be personalised.

The Colour Harmoniser architecture is based upon replaceable models of colour harmony, user interface semantics and a simple characterisation of a user interface. By providing a means to holistically adjust the colours of complete schemes, it enables the rapid generation and exploration of a wide variety of user interface colour schemes. It is also possible that the technique could be extended to other application areas such as fabric design, interior design or environments for computer games.

The research provides experimental evidence for the feasibility of automatically creating user interface colour schemes that are defined in terms of colour relationships, but not hue, and incorporate user-defined design constraints. One possible set of design constraints – distinguishability and grouping – has been defined, and the high scores of the resulting colour schemes indicate that these constraints, while not necessarily optimal, allow the creation of pleasing and usable colour schemes. The improved scores of the adjusted Harmoniser-based schemes illustrate the importance of incorporating human fine-tuning into the colour scheme creation process, and also illustrate that the holistic colour adjustment controls are usable and capable of significantly altering the resulting colour scheme without damaging its integrity.

The experimental results also corroborate the earlier findings of the differences between the genders in colour aesthetics, and broadly validating the saturation/area- and lightness/area-balance models of harmonious colour scheme design. The poor user evaluations of randomly coloured schemes also provides experimental evidence of the paucity of good colour schemes.

The experimental results have also raised some interesting questions: why is there little apparent difference between the quality of the schemes of artists and non-artists? Why are schemes by artists (but not non-artists) and the unadjusted Colour Harmoniser-created schemes penalised by those claiming to have excellent colour sense; and what is the reason for the dip in popularity of unadjusted schemes among viewers in the 55-64 age group? This research has answered some questions and raised others.

Chapter 7

Conclusions

7.1 Summary

The impetus for this research was the observation that the colour selectors in most applications do little to reduce the difficulties in designing good user interface colour schemes. This is a little surprising, as heuristics for the creation of pleasing colour schemes are well-known in the artistic community. Most widely-used applications allow changes to the colour of individual items such as text, shapes and backgrounds, but they provide no support for ensuring that the overall effect is harmonious. In addition to the artistic heuristics, there are also more formal theories that attempt to model harmonious colour appearance, usually in terms of balance, order, area, saturation and lightness. As these guidelines and models are intended to aid artists and graphic designers, factors pertinent to interfaces and usability are not included. Before these theories could be used for the design or evaluation of interface colour schemes, they would need to be extended.

A review of the literature relating to existing work on colour harmony did not uncover any significant extension of the common artistic principles of colour harmony to allow their use in interface design, and most "colour-harmony-in-a-box" tools are based on comparatively simple manipulations of geometrically simple, perceptually non-uniform, usually two-dimensional colour models, and do not take into account information about a particular image, such as the relative areas of its components. A suitable topic for research, therefore, was an investigation into the feasibility of a system that extends the artistic models of colour harmony to operate in a three-dimensional perceptually uniform colour space and includes factors important in user interface design, such as readability and the semantics of interface elements. Such a model would consider an interface colour scheme as a whole, taking account of the relationships between its components during its creation and modification. In this model, the idea of a rotatable wireframe that underlies many of the simpler models of colour harmony has been formalised as the idea of an "abstract colour scheme".

An augmented model of colour harmony based on these ideas was developed and embodied in a piece of software called the Colour Harmoniser, which can analyse the structure of existing Delphi interfaces and create optimised abstract colour schemes.

The augmented model of colour harmony has been incorporated into an evaluation function that can be used to guide the search for schemes whose colour use is harmonious and enhances usability. To allow for subjective colour preference and end-use considerations, users can adjust the resultant schemes holistically to produce a variety of real colour schemes from each optimised abstract scheme. The Colour Harmoniser is a large piece of software, and in terms of the commitment and the type of activities involved, its design and construction constitute a major phase of the project. However, the implementation of the Colour Harmoniser was a scientific exercise, the first of a series of experiments aimed at illustrating the viability of the approach, not an engineering project intended to develop the prototype of a commercial product.

A series of experimental trials was used to validate the augmented model of colour harmony, as exemplified by the Colour Harmoniser's fitness function, and to compare the Colour Harmoniser's algorithmically-created schemes (both before and after user adjustment) with schemes produced using the conventional method of selecting colours (choosing the colours of elements individually).

The results of the trials enabled an assessment of the quality of the colour schemes produced by the two different methods, as well as an assessment of their relative usability.

7.2 The research outcomes

The research has demonstrated that

1. it is possible to design a computationally tractable model of colour harmony that can be used to evaluate user interface colour schemes;
2. it is possible to create harmonious schemes in an abstract colour space;
3. it is possible to build a software tool for creating colour schemes that incorporates the augmented model of colour harmony;
4. users of an algorithmic colour harmony tool do not need colour design expertise;
5. the Colour Harmoniser method produces good colour schemes, and
6. the Colour Harmoniser method of colour adjustment is easy to understand and makes finding good colour schemes easier.

Each of these is now discussed in more detail.

7.2.1 It is possible to design a computationally tractable model of colour harmony that can be used to evaluate user interface colour schemes

This research has sprung from the notion that it should be possible to produce a model of colour harmony that both takes the properties of computer interfaces into account, and can be automated. There are two aspects to this. Firstly, conventional artistic models of colour harmony have had to be extended to accommodate requirements of computer interfaces. Secondly, the augmented model has been expressed in a computationally tractable form, that allows it to be incorporated into a software system, to support the evaluation of competing schemes that are found during an optimisation search process.

Extending colour harmony models for use in interface colour scheme design

Most widely used models of colour harmony are based on geometric relationships within two-dimensional colour wheels, and while such models can be incorporated into algorithms that can produce palettes, they do not take the properties of the items being coloured into account, nor do they produce ready-to-use colour schemes. A notable exception to this general trend is Albert Munsell's three-dimensional model of colour harmony, which is based on simple, though tedious, mathematical operations in a perceptually uniform three-dimensional colour space. However, Munsell's model cannot be used as the basis for an automatic colour selection system, as it is not completely self-consistent; in particular, it is not possible to apply his rule for balancing the colour strength and area of coloured regions to monochromatic or analogous colour schemes, and Munsell does not describe explicitly how to apply the rule to images that contain more than two differently-coloured regions. These problems preclude converting the model, as it stands, into an algorithm. An additional difficulty is that the colour harmony calculations operate on points in the Munsell colour space, which cannot be converted by a simple mathematical function into the colour values that are used for computer displays.

However, a computationally tractable extended version of colour harmony model based on the ideas of Munsell has been produced. The model has been extended to include

- a generalisation of Munsell's two-colour rule to take into account interfaces with an arbitrary number of coloured regions (sec. 4.5.2, p128),
- a way of representing groups of identically coloured elements so that they can be treated as a single region for the purposes of the model (sec. 4.2, p114),
- an extension to the traditional wireframe prototypes to allow for a pragmatic complement to interface aesthetics: large areas of black, white and grey are common in user interfaces, but not allowed for in colour harmony models (sec. 4.4.2, p124). The evaluation of colour strength balance has been modified accordingly (sec. 4.5.2, p128).
- a way of representing the requirement that a given pair of items should be mutually distinguishable (sec. 4.5.5, p131).
- a way of representing the requirement that given text items should be readable against its background. This ensures that interface colour schemes are not made aesthetically harmonious at the expense of readability. (sec. 4.5.6, p132).

- a way of handling monochromatic and analogous colour schemes (sec. 4.5.2, p128).

Algorithmically evaluating interface colour schemes

Using the above model, it is possible to create a fitness function to evaluate the suitability of a particular arrangement of colours for use as an interface colour scheme. The fitness function has two terms related to aesthetics (colour balance and wireframe alignment) and two relating to the colour-related interface semantics (distinguishability and readability) as detailed in section 4.5. The need for each of the four terms has been established by considering the effect of their omission (Table 5.10, p193).

Several experiments established the validity of various aspects of this implementation. The fitness function's evaluation of a colour scheme strongly correlates with that of human evaluators (sec. 5.2.3, p182). The high scores awarded to Harmoniser-based schemes that were created using the evaluations of the fitness function (sec. 6.2.4, p225) also support the contention that the fitness function can serve as a proxy for a human evaluator.

Several factors limit the applicability of the current implementation of the four-term fitness function. Firstly, the evaluation function is specific to the domain of user interfaces. It has only been tested in that context, and, while it may generalise to other domains, this has not been established. However, within this domain, given the high scores of the Harmoniser-based schemes, the use of four criteria contributing to the fitness function is evidently sufficient to allow the reliable algorithmic evaluation of the relationships between coloured items in a user interface colour scheme.

Secondly, the fitness function operates on the assumption that the colour scheme is based on a known wireframe. It can reliably be used when the underlying wireframe is defined (as when creating schemes based on a standard wireframe). However, the validity of its evaluation of pre-existing schemes has not been established, and there may be artistically-created and visually appealing colour schemes that are outside the scope of the simple wireframe/colour balance model.

Thirdly, the fitness function addresses three aspects of colour use within an interface colour scheme (colour aesthetics, readability and distinguishability), but not composition. Human impressions of the quality of visual design are generally holistic and occur subconsciously. If other aspects of a design, particularly its composition, are poor, it may be condemned as a whole, colour scheme and all.

Fourthly, the fitness function does not assess hue choice. Human fine-tuning of an automatically created abstract colour scheme by rotation to a particular hue orientation can distinctly improve the visual appeal of a colour scheme, as is evident from the higher scores of the adjusted schemes when compared to the unadjusted schemes (sec. 6.2.4, p225). Fine tuning can affect the aesthetic appeal of an interface, but does so without changing the chromatic relationships between the interface elements, and therefore without altering the fitness function score.

In summary, the fitness function is not complete. It can be used to gain a broad holistic assessment of an interface colour scheme, but not to gauge subtle effects, such as those related to the use of particular colours. For these, human input is highly desirable.

7.2.2 It is possible to create harmonious schemes in an abstract colour space

Algorithmic creation of colour schemes is straightforward, but the creation of good colour schemes is not. This is clear from the very poor ratings given to schemes with randomly coloured elements in the "Compare the Results" experiment (sec. 6.2.4, p225). Therefore, to ensure that an algorithmic method produces satisfactory interface colour schemes, it must include a means of assessing the colour aesthetics, a means of ensuring that usability is not impaired by poor colour choice, and a means of allowing for the pragmatic and subjective nature of human colour preference.

To address these concerns, the creation of a colour scheme is broken into two stages: establishing a set of chromatic relationships between the interface components in an abstract colour space, without specifying any actual colours; and transforming the abstract colour scheme into real colours that reflect the preferences of the human developer.

The first stage – designing colour scheme in terms of relationships – is worthwhile because the harmony of a colour scheme is widely believed to derive from the chromatic relationships between coloured elements, independent of the actual hues involved. This is evident from the artistic colour harmony models discussed in section 2.9 (p52) and the extended colour harmony model (sec. 4.5, p126), both of which are parameterised in terms of the elements being coloured and their relationships in colour space, not in terms of any real colours. This independence of colour schemes from hue allows schemes to be defined as geometric arrangements in a perceptually uniform colour space, as colour molecules (sec. 4.4, p123) that can be scaled and rotated within the space without affecting the harmony of the colour scheme, and without fixing on any particular hues.

In order to find abstract "colourings" that satisfy both the colour harmony and usability aspects, a constraint satisfaction process is used. This optimises the placement of items in the abstract space to ensure that the resulting scheme (when mapped to real colours) will be harmonious and readable, and that pairs of items will (where required) have distinct colourings. Satisfying these conditions requires compromises. Two of the terms (readability and distinguishability) are important to usability and are given priority during the constraint satisfaction, whereas the aesthetic terms may be compromised if necessary. No single arrangement of colours can be guaranteed to be aesthetically pleasing (there are factors not assessed by the fitness function), so the optimisation process produces not one, but several, schemes for the user's consideration.

The second stage – incorporating human intent into a colour scheme – can be broken into a constraints definition phase (prior to scheme creation) in which the designer specifies the required colouring relationships between the elements, and a fine-tuning phase (after the tool has produced a conformant colour molecule) in which holistic alterations are made, according to the user's judgement. The latter phase is virtually impossible to quantify or define via specifications. Colloquially, it is one of those judgements that seem common in design: "I'll know it when I see it". However, while the characteristics of such colour schemes cannot be precisely specified, to facilitate their discovery, the Colour Harmoniser prototype provides a means of interactive colour scheme exploration. A set of high scoring schemes is presented at the end of the colour scheme creation phase (sec. 4.6.3, p142). From these, the user can select a scheme whose colour composition appeals and then, using the personalisation controls, explore a wide range of colourings by adjusting the hues used, their saturation and which elements are light and dark (sec. 4.8, p156). It was for these reasons that the decision was made to design the colour harmonisation tool so that it requires two distinct types of input from the user: first, preferences on the overall structure of the colour scheme, and later, the interactive selection of the colours to be used.

At the start of the project, it was intended that the colour atoms should be manipulated in a perceptually uniform abstract colour space, and then mapped to sRGB colours for display. The first part of this plan was followed, and optimisation within the abstract space functions as expected. However, there is a problem concerning the mismatch between the idealised spherical shape of the abstract colour space and the varying distances allowable within the perceptually uniform colour spaces if the colours are to be displayable on monitors[1]. Designing the abstract colour schemes in a real perceptually uniform space (e.g. CIELAB) would have been problematic due to its irregular shape, so an idealised perceptually uniform space – the abstract colour space – was used instead. The mapping of the colour scheme from this abstract space to a real colour space was deferred until after a scheme had been designed. This mapping requires compromises, as the colour molecules that will fit completely within the sRGB gamut boundary in the CIELAB space, at all orientations, are severely limited in saturation. Instead, the colour molecule is scaled to give well saturated colour schemes over a wide range of hues, although the scaled colour molecule will not fit within the gamut boundary at all orientations (sec 4.7.1, p149 and sec. 4.7.2, p154).

The method selected for handling out-of-gamut colours in the Colour Harmoniser prototype was to allow them to map to nearby colours. This can cause some of the displayed colours to vary from those calculated, but the effect is not normally apparent. This is clear from the results of the comparative colour scheme creation experiment, whose participants found the effects of the controls beneficial and easy to understand (sec. 6.1.5, p216), even though, via the colour strength control, they could force some colours out-of-gamut. Doing so will subtly alter the hues and the distinguishability, but the effect is gradual and, in practice, does not appear to be problematic.

In the Colour Harmoniser method, colour schemes are optimised in an abstract colour space that, like the CIELAB colour space, is perceptually uniform. Lightness and saturation in the abstract space can be mapped to lightness and saturation in the CIELAB space, but, as no mapping to actual hues is defined (sec. 4.7, p147), users cannot experience the abstract colour scheme directly. Therefore, to display an abstract colour scheme to the user for fine-tuning, an initial orientation must be chosen. The wireframe rotation angle (governing the hue) can be predetermined, or it can be chosen

[1] The limitation is more severe with display gamuts, but also exists with the range of perceivable colours.

randomly. Which option is chosen is not particularly important since the rotation angle is immediately changeable by the designer using the personalisation controls. While not part of this research, it may be possible to preorient the wireframe so the initially displayed schemes use colours that are appropriate for the intended users. This would require the user group to be defined as part of the initial constraints, and would also require a method of finding preferred colourings for particular user groups, using for example, the colour suitability database of Morton and Peterson (2007), or the palettes of Kobayashi (1981). However, for best results, human fine-tuning would still be desirable.

7.2.3 It is possible to build a software tool that incorporates the augmented model of colour harmony

The contention that it is feasible to automatically produce acceptable interface colour schemes is fundamental to this project, and an algorithmic evaluation of colour schemes described in the previous section is clearly a move towards validating this position. The next step was to demonstrate the feasibility of producing a software artifact that could successfully produce conformant colour schemes (that is, schemes that conform to the definition of colour harmony described by the model, whether or not any humans agree). The design of this proof-of-concept system, the Colour Harmoniser, was described in Chapter 3, and its implementation was described in Chapter 4.

It should be emphasised that the development of the Colour Harmoniser was not the point of the research; it is merely a vehicle for demonstrating, by example, the possibility of creating an automatic tool. Consequently, the exact structure of the software and the design of its component parts are less important than the general conclusion that such a tool can be produced. Nevertheless, as the Colour Harmoniser uses a novel approach, the derivation of the major architectural components deserve recognition as components of the validation.

The major components within the architecture that have enabled the Colour Harmoniser to create harmonious colour schemes are:

- a mechanism for defining the chromatic relationships between selected interface elements (sec. 4.2, p114),

- a mechanism for grouping interface components that have been created as separate components by the development software but which, for the purposes of the colour harmony model, need to be treated as a single item (sec. 4.2, p114).

- a genetic optimiser that facilitates searching the solution space for colour schemes that conform to the colour harmony and usability constraints for a particular interface (sec. 4.6.2, p139).

- an interactive method of holistically adjusting colour schemes that is both readily comprehensible and effective, as will be summarised in section 7.2.6.

- a fitness, or objective, function that rates colour schemes using the extended model, and is used by the genetic optimiser to direct the search for harmonious colour schemes. (sec. 4.5, p126).

  This function includes terms for

  - evaluating the colour aesthetics, as determined by the balance of colour strength (sec. 4.5.2, p128) and the alignment with colours from prototypical colour scheme paths (a wireframe) in colour space (sec. 4.5.4, p130).
  - evaluating readability as part of the overall evaluation of a interface colour scheme (sec. 4.5.6, p132).
  - evaluating the distinguishability of specified pairs of elements (sec. 4.5.5, p131).
  - reducing the effect of large achromatic areas on colour balance when the black-white axis is added to the wireframe (sec. 4.5.3, p130).

Subtle internal interactions between the elements of a colour scheme affect its overall appeal. Therefore, to ensure that the colour schemes produced by the Colour Harmoniser method are also good colour schemes, the corroboration of human evaluators is required.

There are two forms of user input into the colour scheme design process: the pre-creation specification of colouring constraints prior to scheme creation, and the post-creation holistic selection of the colours appears to work well, as is evident from the high score of the Harmoniser-adjusted schemes created using this method. However, the initial constraints of grouping and distinguishability, while effective, are not particularly subtle: groups can be used to colour elements identically, and defining a distinguishability constraint between elements causes them to have as much colour contrast as possible. A method allowing a more graduated set of colouring relationships is suggested in section 7.5. Likewise, the holistic colouring controls are effective, but change all the colours, which may not be desired. A method that allows individual adjustment, something that is precluded in the current implementation, is also discussed in section 7.5.

Currently, the user is required to specify the wireframe to be used as the basis for the colour scheme, but this could be omitted. Instead, the Harmoniser could create a small number of schemes for each of several wireframes (e.g. two for each of five wireframes) to form a set of ten schemes offered at the end of the optimisation. While the current implementation shows colour scheme variations based on the same wireframe, there is no technical reason why schemes from several different wireframe could not be shown instead. Which option is preferable would depend on whether a variety of colourings were wanted or whether it was desirable to specify the number of hues used in the colour scheme (and thereby determining the wireframe).

Usability constraints together with area, order, and known colour space distances, are sufficient parameters for colour scheme design

It was found that, when augmented with semantic constraints, the use of area, order, and distances in colour space (including a known lightness axis) are sufficient to allow the evaluation of user interface colour schemes, subject to the conditions noted earlier in this section. This is established by demonstration. These inputs are the only parameters used in the evaluation of colour schemes during their initial creation by the Colour Harmoniser prototype. The sufficiency is demonstrated by the high scores of Colour Harmoniser derived schemes (both raw and adjusted) in the "Compare the Results" experiment (sec. 6.2.4, p228).

7.2.4 Users of an algorithmic colour harmony tool do not need colour design expertise

One of original aims of the research project was to aid developers unskilled in colour design to create colour schemes for user interfaces. Although some educational institutions combine tuition in graphic design and software design within one programme of study, the fields of software engineering and web site development are full of people whose skills are restricted to the technical aspects, and there is no reason to suspect that this situation will change rapidly, if at all.

This is the reason why the approach to interface design that is described in this thesis is likely to be of practical use: there seems to be a genuine need for a tool that can perform complex technical calculations to compensate for designers' lack of colour design skill, which are all too often replaced by arbitrary preferences for colours that do not harmonise, or the safe "no choice" option of using default colours, which are often rather bland.

In the first phase of colour scheme creation using the Colour Harmoniser method, users do have to input some information about the interface: specifying groups of elements which have identical semantics within the user interface; identifying element-pairs whose colours must be distinct; and selecting an underlying colour scheme (the wireframe) (sec. 4.2, p114). These inputs do require decisions to be made about the interface, but none of them are related to colour harmony; in principle at least, they are things that a system designer will already know about and will not require significant thought. No judgements are required about whether or not colour schemes conform to colour harmony criteria.

For clarity, in the preceding chapters, the underlying colour scheme has been referred to using its artistic name (e.g. complementary, split-complementary, monochromatic, etc.), but there is no necessity for these design-oriented terms. For a non-technical audience, more descriptive terms indicating the number of colours in the final scheme could be used: e.g. a single colour scheme (monochromatic), two colours (complementary), three colours (split-complementary), related colours (analogous) etc. The *"Include Black-White"* option could be relabelled in a similar fashion.

The second stage of colour scheme creation (creating and optimising colour molecules) is automatic, and the third stage (colour scheme selection and personalisation), as has been demonstrated in the "Compare the Methods" experiment, is non-technical and readily comprehensible.

It is therefore clear that no knowledge of colour design theory or vocabulary is required to specify, create, or fine-tune colour schemes using the Colour Harmoniser method.

7.2.5 The Colour Harmoniser method produces good colour schemes

There would be little point in developing a system for automatically deriving harmonious colour schemes if it did not produce good schemes. Chapter 6 documented the experiments that were conducted with the overall aim of determining where, on some notional scale of colour scheme quality, the colour schemes produced by the Colour Harmoniser would fall. It was expected that randomly coloured schemes would fall at the bottom of the scale, that schemes produced by human non-artists would fall at an intermediate position and that schemes produced by human artists would be at the top of the scale. It was hoped that the schemes produced by the Colour Harmoniser would approach, or possibly improve slightly, on those produced by human non-artists.

The results of the pilot fitness function validation experiment (sec. 5.7, p181) and the main fitness function validation experiment (Table 5.2.3, p185) clearly demonstrate the high correlation of the fitness function's assessment of a colour scheme with those of human assessors. These results indicate that the human and fitness function assessments agree, but do not allow an assessment of the overall quality of the Harmoniser-based schemes. This was evaluated in the final "Compare the Results" experiment in which the participants scored Harmoniser-based schemes along with human-created and randomly coloured schemes, without being aware of how the schemes were created. The results were unequivocally in favour of the schemes produced by the Colour Harmoniser.

Harmoniser-created colour schemes, before adjustment, are comparable to those by human designers

The Harmoniser-based schemes incorporate what have long been thought to be the critical factors for artistic colour harmony, augmented to ensure that usability and semantic aspects are included. Broadly speaking therefore, the unadjusted Harmoniser-created schemes should be acceptable. This was found to be correct, as the participants awarded the unadjusted Harmoniser-based schemes mean scores that were equal to or higher than the scores awarded to schemes created by both artists and non-artists (Table 6.2.4, p225), across all four evaluation criteria (visual appeal, professionalism, how artistic and website suitability). These results are detailed in the results of the "Compare the Results" experiment (sec. 6.2.4, p229). The high scores of the unadjusted schemes used in this trial can be seen as validating the choice of the factors incorporated into the fitness function.

Harmoniser-based schemes, after adjustment, are better than those of many human designers

The adjusted Harmoniser-created schemes were scored more highly (Table 6.2.4, p225) than schemes created using any of the other four methods (random colouring, by non-artist, by artists, and by the unadjusted Harmoniser-based schemes) in three of the four evaluation criteria. The differences are statistically significant (Table 6.4, p228). All Harmoniser-created schemes conform to the heuristics for artistic colour harmony. From a pool of schemes having these characteristics, the best have been selected and then fine-tuned by the human participants[2] who can choose colours to optimise what the fitness function cannot assess, the subtle nuances of colour and their interactions. This is clearly the case, as the scores after human adjustment are significant higher than the scores before the adjustment (those awarded to the Harmoniser-raw schemes).

Harmoniser-based schemes have broad appeal

Across multiple demographic categories (gender, age, and artistic and colour design backgrounds), the adjusted Harmoniser-based colour schemes scored significantly more highly than the schemes by artists, non-artists and the Colour Harmoniser (when unadjusted) for three of the four evaluation criteria (visual appeal, professionalism, and website suitability). For the fourth criterion (how artistic), the adjusted Harmoniser-based schemes were scored more highly than those by non-artists, and were comparable to the schemes by artists and unadjusted schemes from the Colour Harmoniser prototype.

The participants' self-assessment was the basis for the artist vs. non-artist categorisation. Self-assessments are subject to bias in both directions (both under- and over-rating expertise). It is not clear whether the results of the "Compare the Results" experiment would differ if the schemes being assessed included some from people whose expertise is demonstrated rather than based solely on self-assessment. Such schemes could come from (for example) practicing artists and professional design-oriented (not software-oriented) web developers. These schemes could be included along with those from participants with no background in colour design, and randomly coloured schemes. If the demographic survey retained the question on the participant's self-assessed "artist" categorisation, this would allow the analysis of the results to determine whether, for example, the evaluations of professional developers and practising artists vary significantly from those of the rest of the participants, and also whether the evaluations of those with demonstrated expertise differs significantly from those whose expertise is based solely on self-assessment.

The experimental trials used an interface that is typical of a wide range of interfaces, both those of desktop and online applications, as discussed in section 5.1.1 (p167). Now that it is clear that the Colour Harmoniser method can produce good colour schemes for such a generic interface, further experimental trials to confirm its applicability to a wider range of interface styles are warranted.

7.2.6 The Colour Harmoniser method of colour adjustment is easy to understand and makes finding good colour schemes easier

A number of questions in the user surveys were designed to assess the users' impressions of the usability of the Colour Harmoniser method of colour scheme adjustment. In some

[2] the selection and user-adjustment of the raw Harmoniser-based images was done during the "Compare the Methods" experiment

respects, the results of such questions are inherently unreliable: as the tool is the first of its kind, the users have little basis for comparison, and it is therefore difficult to formulate questions that separate usability from functionality. Perhaps, since this is the first such tool, that is not of the utmost importance; easily accessible functionality that provides useful facilities not previously available is a major, if not the major, contribution to usability.

The method of allowing users to find a Harmoniser-based scheme and adjust it to their liking was evaluated during the "Compare the Methods" experiment. The participants indicated that

- *the holistic adjustment of colour schemes is easy to understand:* this is evident from the responses to related questions from the "Compare the Methods" experiment.

- *the holistic controls are effective:* this is evident from responses to related questions from the "Compare the Methods" experiment and the higher scores given to adjusted Harmoniser schemes in the "Compare the Results" experiment also support this conclusion..

- *the automatically created schemes can be improved by simple user adjustment:* both the survey responses from the "Compare the Methods" experiment and the higher scores given to adjusted Harmoniser-based schemes in the "Compare the Results" experiment support this conclusion also.

- *holistic methods of colour scheme selection are less frustrating:* from the results of the "Compare the Methods" experiment.

- *colour scheme creation times using the Harmoniser and conventional techniques are comparable:* there was no significant difference between the time taken to find a scheme using either method in the"Compare the Methods" experiment. However, it is possible to explore a much larger number of colourings using the Harmoniser-based method, as altering any of the direct-manipulation personalisation controls will recolour the complete interface. Achieving the same effect manually would require changing the colour of each element individually (possibly with multiple attempts to get the desired colour), which could lead to a reluctance to vary the colours greatly from those initially selected. Therefore, although it was hoped that the Harmoniser-based method would be faster, it appears that the users spent similar amounts of time whatever method was used. However, as detailed in the next point, the results were not equivalent.

- *finding professional schemes is expected to be faster using the Harmoniser method:*

  - the participants strongly indicated they expected the Harmoniser method would enable a professional-looking scheme to be found more quickly ("Compare the Methods" experiment results).

– the results using the Harmoniser method, after adjustment, were judged to be better in several criteria (including "professionalism") than the manually created schemes, by both artists and non-artists ("Compare the Results" experiment). Combining this result with the fact that the participants spent similar amounts of time creating schemes using the manual or the Harmoniser-based method (sec. 6.1.5, p214), provides indirect confirmation of the users' expectation – that professional looking schemes would be found more quickly using the Harmoniser-based methods – as, for schemes that took similar amounts of time to prepare, those from the Harmoniser-based method are judged as having a more professional appearance.

The user evaluations described above related only to the post-harmonisation phase. The earlier data-capture phase involves a complex interface that was designed as an interim measure – a developmental prototype – not as part of a realistic system. The development of an appropriate interface for the grouping and definition of distinguishability constraints is a different project from establishing the viability of the overall Colour Harmoniser method. Some ideas relating to a simpler interface are included in the section 7.5).

7.3 Confirmation of the thesis

The original stated aim of the research was stated in section 1.3 (p8) as:

> *The system would initially define the relationships between colours in an abstract colour space, and use these to create a family of raw, or unadjusted, colour schemes, each of which could subsequently be mapped to many sets of real colours. This mapping could be accomplished by incorporating a simple mechanism to allow users to holistically tweak the colouring to their preference. This research would be considered successful if users considered the quality of the raw schemes to be comparable to the quality of schemes created by human developers unskilled in colour design, and if users considered the final (tweaked) schemes to be of significantly higher quality than the raw schemes.*
>
> *To validate this thesis, it is necessary to*
>
> - *extend the artistic models of colour harmony so that they are:*
>   - *applicable to user interface design,*
>   - *susceptible to being described and used algorithmically,*
>   - *incorporate user preferences with regard to GUI characteristics and overall colouring.*
> - *develop a software tool that incorporates the extended colour harmony model, and can be used to create user interface colour schemes and facilitate their adjustment.*
> - *experimentally validate the extended colour harmony model through user evaluations of the resulting colour schemes with those created using a more conventional method of colour selection.*

- *test whether users find the holistic method of colour selection and adjustment easier to use than conventional methods.*

As discussed in the research outcomes, all of the above goals have been accomplished. It is therefore possible to conclude that the original thesis has been shown to be well supported.

7.4 Contributions to knowledge

1. An extended model of colour harmony that allows for a richer image than classical models, accounts for more aspects of colour use than most conventional models, and can be used for the design of computer interfaces.

2. A metric for assessing colour scheme quality that can be applied to computer interfaces and can be implemented as an automatic process.

3. An approach to creating usable and harmonious colour schemes for user interfaces that allows humans to express their taste, but does not require knowledge of colour theory or design.

4. An abstract representation of colour schemes that allows the colours to be changed holistically without destroying the harmony of the colour scheme.

7.5 Possible extensions

Several topics not central to this thesis have been noted previously but not explored. These include extensions to the model of colour harmony and extensions to the holistic method of creating and adjusting colour schemes.

7.5.1 Extending the model of user interface colour harmony

The current extended colour harmony model detailed in section 4.5 (p126) and the subsequent use of this model in the Colour Harmoniser prototype is sufficiently general to model the colour schemes found in a wide variety of user interfaces. However, there are additional features that could be added to the colour harmony model and to the fitness function, and to a tool for creating user interface colour schemes. These extensions include

interfaces that include texture, gradation and transparency:
The solidly coloured elements used in the current project exclude both those with texture and those with gradation.

The methods used to integrate textures would depend on whether the colour of the textured element can be updated by the Colour Harmoniser method or not. If so, then a single textured element could be represented as multiple pseudo-elements that are positioned in the abstract colour space to form a dithering palette. The total area of the pseudo-elements would be the same as the original element, and the colour relationships between them within the abstract colour space would allow the colouring in the original texture to be (approximately) recreated. These pseudo-elements would be evaluated as part of the colour scheme during the optimisation, but their relative positions would not be altered by the optimiser: to allow the texture to be recreated, their colour relationships must be preserved. The optimiser could however, move all the textured pseudo-elements (as a set) and move the other colour atom as usual. If the colours of the textured element were fixed (and not updateable), the methods described in the next section (sec. 7.5.2) on the inclusion of fixed coloured element may more relevant.

Elements with gradations would need the model to include the geometry and the direction or shape of the gradation to be included in the calculation. The presence of semi-transparent elements would introduce an even more serious complication, as the colour of any point within the interface could be a function of all the elements at that point, depending on the transparency of each, and their depth ordering. The effect of blending the colours from semi-transparent objects is considered by Tornquist (1999), but how such effects could be incorporated into a colour harmony model, and into the architecture of a system for automatic interface colouring, has yet to be determined.

maintaining colour harmony over multiple pages:
User interfaces can have multiple views, and websites normally consist of multiple pages. The need to maintain harmony over multiple pages has not been considered, but the need for coherent colouring would be a paramount concern if the system was extended to handle multiple pages: simply applying the current model to multiple pages is unlikely to be satisfactory, as differing areas of elements on different pages would result in variations of lightness and saturation of colours between pages. From a user perspective, the changes would appear arbitrary, would be perplexing, and are best avoided.

groups with independently specified colour schemes:
In the architecture as described, the only option for the colouring of grouped elements is for them to all be the same colour. Instead, grouped elements could have a separate (smaller) independently selected wireframe, which would allow each group to have its own local colour scheme. Such a facility was incorporated into an early version of the Colour Harmoniser prototype (Moretti and Lyons, 2005), but development was discontinued when it was noted that local schemes could result in elements forming part of a local colour scheme appearing to be haphazardly coloured and detracting from the overall appearance. It is not that sub-schemes are unworkable – they could be used to provide local colour variation to add interest – but additional constraints would be needed to ensure that the local variation does not damage the appearance of a cohesive and ordered colour scheme. The ability of a local colour scheme to satisfy distinguishability constraints would also be limited, as, necessarily, a local scheme can use only a subset of the colour space.

additional wireframe orientations:

Several of the current wireframes are tilted at 45° to the horizontal and are flat: wireframes such as the split-complementary have both ends of the Y at the same vertical position, and the elliptical scheme has both sides at the same vertical position. Additional colouring options could be obtained by rotating such wireframes along their long axis. For the split-complementary scheme (for example), this would tilt the wireframe so that one end of the Y was higher than the other. Another option would be to allow any of the complementary, split-complementary and elliptical wireframes to tilt so that they were not at 45° , but were more (or less) vertical. Varying the wireframe orientation could affect the readability of elements, and so might need to be taken into account during the optimisation. However, these variations do indicate that, without extending the Colour Harmoniser architecture, there are options for personalising colour schemes beyond simply varying their hue, saturation, and the lightness of elements.

viewing groupings and distinguishability constraints holistically:

There are many different interfaces that could be used to allow interface elements to be grouped and have distinguishability constraints defined between them. It has been assumed that these concepts of grouping and distinguishability would be easily understood by developers unskilled in colour design. However, this assumption has not be validated by user trials.

The hierarchical tree-structured grouping interface used to characterise an interface in the Colour Harmoniser prototype enables the appropriate functionality, but only allows the display of the distinguishability of one element at a time (see fig. 4.5, p119). More holistic view of constraints and groups are possible. For example, an interface could initially list the names of all the elements in a panel to the left of a canvas. These would be all those elements whose grouping and distinguishability constraints are undefined. Any of these element names could be dragged onto the canvas and grouped into clusters. These clusters (which would be the groups) could themselves be dragged on the canvas (to space the groups apart). To define a relationship between clusters, a connecting line would be drawn from one cluster to another. Differing colours or line-types could be used to indicate different relationships or constraints. The proximity of one cluster to another could be used be used to indicate a different type of relationship, for example, similar colouring: a less forceful constraint than distinguishability. This canvas-based interface would allow the definition of groups and distinguishability constraints to be defined, while providing an overview of all the constraints simultaneously.

An enhancement would be the ability to select an element, not from a list of names, but by dragging an element of the interface being coloured onto the canvas, which would drop its name at the mouse-up location.

limiting readability and distinguishability contrast:

The current definitions of the readability and distinguishability fitness function terms return higher values of fitness for increasing lightness and chromatic contrast. The optimiser, in maximising the fitness score, will move items further apart than is necessary for readable text and distinguishable elements. It may be desirable to ensure that the items have sufficient contrast rather than as much as possible, as this would result in more subtle colourings and would allow more flexibility in the constraint resolution.

intermediate distinguishability colouring constraints:
The "is-distinct-from" distinguishability constraints allows essential distinctions between elements to be captured and reflected in the colouring. However, the distinctions are not subtle. The introduction of a "is-similar-to" chromatic relationship would allow an intermediate option. This would require there to be a peak in the distinguishability fitness at the desired separation in colour space, and a falling fitness as the separation becomes too great for the items to appear similar. This is a more flexible variation of the "local colour scheme" mentioned above, but, would also require care to be taken to ensure differences in colours of the elements with the "is-similar-to" constraint fit visually within the overall colour scheme.

7.5.2 Extensions to the holistic creation and adjustment of colour schemes

The current architecture is intended to minimise the requirement for up-front specification and allow pleasing colour schemes to be discovered by the user by holistically adjusting the final scheme colouring. There are variations on this theme that allow for differing degrees of flexibility in the alteration of colours and the need to specify constraints.

Incorporating fixed colours into automatically-created colour schemes

The current model assumes that the wireframe may be freely rotated, but does not allow fixed colours (e.g. a logo) to be incorporated in the final colour scheme. Fixed colours correspond to fixed points in the sRGB, CIELAB and abstract colour spaces. It may be possible to find a wireframe and a rotation that would result in a fixed point in abstract colour space being on the wireframe or, if this is not possible, the wireframe rotated so that one of its segments uses the same hue as the fixed point, e.g. a split-complementary Y-shaped wireframe allows three rotations that would align a hue with a single fixed colour. Fixed colouring does remove the freedom to arbitrarily rotate the wireframe, but given the large number of orderings on a wireframe, this would still allow a variety of colour schemes, all of which harmonise with the required colour.

If there are multiple colours in the item to be added (e.g. a photograph), it may not be possible (or desirable) to force all the colours to be near the wireframe. However, the dominant one or two hues could be determined and the wireframe oriented so the overall scheme harmonises with those, accepting that there may be some colours within the photo that don't harmonise. Using the dominant hue or hues will ensure that, overall, the most used colours in the photo fit with the rest of the interface colour scheme, and if the photograph had a definite edges (as is common), these would act to visually contain any colours that do not harmonise.

Allowing the non-holistic adjustment of colours elements within a scheme

The automatic creation of colour molecules results in colour schemes that are broadly acceptable, and the holistic controls allow all the colours to be adjusted. However, there may be occasions when a user may wish to adjust the colour of a single interface element without changing any of the others. Currently, this is not possible. However, during the fine-tuning stage, it would be possible to allow the user to freely alter any element colour, possibly with oversight by a module using the fitness function to inform the user if the changes are severely detrimental. The designer would be unlikely to impair the usability of the scheme, but if they have limited aesthetic judgement, they may detrimentally change the colour aesthetics without being aware of having done so. To limit such detrimental effects, changes to the colours of individual elements could be restricted to altering their lightness and saturation, but not their hue.

Enabling end-user personalisation of colour schemes

This research has focussed on the creation of interface colour schemes from the perspective of an application or web site developer, but their taste in colours may differ from that of an end-user. The scheme creation and personalisation phases of the Harmoniser method are essentially independent: abstract colour schemes can be exported as XML, then imported into other software and used to adjust a colour scheme without requiring the software modules originally used by the developer to create the abstract colour schemes. Therefore, the colour personalisation (rotation/saturation/flip light-dark) controls could be made available to an end-user to tailor an interface colour scheme to their liking and save the setting as a personalised option. The participants in the "Compare the Methods" experiment have used a similar method of selection and adjustment when fine-tuning the pre-created Harmoniser-based schemes, and have indicated that the controls were easy to understand and use (sec. 6.1.5, p216).

Adding this additional stage of personalisation to an end-user application or web site would be computationally inexpensive, and would allow an end-user to select from a very large number of colour schemes, all reflecting the colour relationships as specified by the developer, but using colours preferred by the end-user.

7.6 General conclusions

In a narrow sense, this research has clearly demonstrated that it is possible to automate the production of colour schemes for computer interfaces. The process that has been developed is based on the classical artistic rules of colour harmony, a well-known, but informal and incomplete, set of rules of thumb. Completing that set of rules, in accordance with (and extending) Munsell's suggestions, and converting them into a colour scheme quality metric, has led to the development of a more successful system than the researchers ever dared to hope.

From a wider perspective, however, this research hints at the possibility of mechanising other aspects of visual aesthetics. For example, it may be possible to extend the colour harmony model that has been developed in this research to allow for textured and gradated interface components. Colour harmony is not the only field where existing informal rules of thumb may be susceptible to formalisation and automation. For example, image composition – the disposition of shapes on a two-dimensional surface – is often described in terms of well-known, but vaguely expressed, rules that may also be susceptible to formalisation. One could perhaps develop software that prompts developers to place important items at the golden mean of the image, or uses the "rule-of-thirds" to position objects in visually interesting ways, to help them to balance areas visually (as well as chromatically), or use tapering diagonals to lead the gaze into the image, and prevent it from sliding out. Image composition is a richer field than colour choice, and the standard compositional guidelines may be too contradictory to allow the degree of automation that has been possible in the current development, but it would be an interesting project, and clearly, such a development in the area of layout would complement the chromatic initiatives described here.

The colour selection techniques that have been developed for user interfaces may be adaptable to other problem areas: the visible surfaces in our houses – walls, floors, windows and even ceilings – are adorned with wallpaper, paint, carpet, and curtains whose colours can be chosen freely. Whereas many colours were once restricted to the wealthy, today, this is no longer an issue. Colours may be freely chosen, and pigments are more vibrant than ever before. However, these can result in colour choices that are later regretted. While poor choices may be temporarily irritating with clothing, they are both more expensive and harder to change with interior design. However, this may well change. With the ever-increasing pace of technological development, large flat panel displays are now commonplace, enabling electronic "windows" that were pure fantasy not long ago. Currently in development and the early stages of production are a variety of technologies for flexible colour displays and coloured electronic paper. These raise the interesting possibility of dynamically changing the colouring of objects in our living environment. While not yet here, the idea of electronically recolouring a wall is no longer restricted to the realm of science fiction. When such technology becomes available, the idea of dynamically changing a room's colour scheme using the personalisation controls is an intriguing one.

We find ourselves searching for colour harmony more often than our ancestors, but are no better equipped to find it than they were. The ideas explored in the current research are thus apposite and timely. They may foreshadow the development of a range of tools that, like most technologies, are used in areas and in ways that were never envisaged by their developers.

Appendix A

Expanded results table for the "Compare the Results" experiment

A.0.1 Cross-tabulation of the t-test results of the difference in means for differing colour scheme creation methods

Visual Appeal

	CH-Adj	CH-Raw	Artist	Non-artist	Random
CH-Adj	–	t=3.0 df=712.4 p=0.003 **	t=3.9 df=694.7 p<0.001 ***	t=4.2 df=685.4 p<0.001 ***	t=19.3 df=711.1 p<0.001 ***
CH-Raw	t=-3.0 df=712.4 p=0.003 **	–	t=1.0 df=713.2 p=0.316	t=1.3 df=704.8 p=0.194	t=15.2 df=699.1 p<0.001 ***
Artist	t=-3.9 df=694.7 p<0.001 ***	t=-1.0 df=713.2 p=0.316	–	t=0.3 df=716.9 p=0.778	t=13.2 df=677.7 p<0.001 ***
Non-artist	t=-4.2 df=685.4 p<0.001 ***	t=-1.3 df=704.8 p=0.194	t=-0.3 df=716.9 p=0.778	–	t=12.9 df=668.2 p<0.001 ***
Random	t=-19.3 df=711.1 p<0.001 ***	t=-15.2 df=699.1 p<0.001 ***	t=-13.2 df=677.7 p<0.001 ***	t=-12.9 df=668.2 p<0.001 ***	–

Professionalism

	CH-Adj	CH-Raw	Artist	Non-artist	Random
CH-Adj	–	t=3.8 df=716.6 p<0.001 ***	t=5.1 df=712.5 p<0.001 ***	t=5.8 df=703.5 p<0.001 ***	t=19.8 df=697.2 p<0.001 ***
CH-Raw	t=-3.8 df=716.6 p<0.001 ***	–	t=1.4 df=715.6 p=0.159	t=2.1 df=707.3 p=0.041 *	t=15.3 df=685.7 p<0.001 ***
Artist	t=-5.1 df=712.5 p<0.001 ***	t=-1.4 df=715.6 p=0.159	–	t=0.6 df=712.9 p=0.533	t=13.2 df=672.5 p<0.001 ***
Non-artist	t=-5.8 df=703.5 p<0.001 ***	t=-2.1 df=707.3 p=0.041 *	t=-0.6 df=712.9 p=0.533	–	t=12.4 df=661.7 p<0.001 ***
Random	t=-19.8 df=697.2 p<0.001 ***	t=-15.3 df=685.7 p<0.001 ***	t=-13.2 df=672.5 p<0.001 ***	t=-12.4 df=661.7 p<0.001 ***	–

Artistic appearance

	CH-Adj	CH-Raw	Artist	Non-artist	Random
CH-Adj	–	t=1.6 df=695.7 p=0.121	t=1.6 df=687.8 p=0.121	t=2.9 df=678.0 p=0.004 **	t=14.4 df=696.0 p<0.001 ***
CH-Raw	t=-1.6 df=695.7 p=0.121	–	t=0.1 df=703.7 p=0.950	t=1.3 df=695.5 p=0.184	t=12.1 df=693.9 p<0.001 ***
Artist	t=-1.6 df=687.8 p=0.121	t=-0.1 df=703.7 p=0.950	–	t=1.2 df=704.9 p=0.222	t=11.5 df=690.6 p<0.001 ***
Non-artist	t=-2.9 df=678.0 p=0.004 **	t=-1.3 df=695.5 p=0.184	t=-1.2 df=704.9 p=0.222	–	t=10.2 df=681.5 p<0.001 ***
Random	t=-14.4 df=696.0 p<0.001 ***	t=-12.1 df=693.9 p<0.001 ***	t=-11.5 df=690.6 p<0.001 ***	t=-10.2 df=681.5 p<0.001 ***	–

Web-site Suitability

	CH-Adj	CH-Raw	Artist	Non-artist	Random
CH-Adj	–	t=3.7 df=710.2 p<0.001 ***	t=4.2 df=711.8 p<0.001 ***	t=5.4 df=698.8 p<0.001 ***	t=21.3 df=669.4 p<0.001 ***
CH-Raw	t=-3.7 df=710.2 p<0.001 ***	–	t=0.5 df=717.1 p=0.607	t=1.8 df=707.0 p=0.076 .	t=16.2 df=644.4 p<0.001 ***
Artist	t=-4.2 df=711.8 p<0.001 ***	t=-0.5 df=717.1 p=0.607	–	t=1.2 df=715.3 p=0.214	t=15.2 df=638.5 p<0.001 ***
Non-artist	t=-5.4 df=698.8 p<0.001 ***	t=-1.8 df=707.0 p=0.076 .	t=-1.2 df=715.3 p=0.214	–	t=13.5 df=620.3 p<0.001 ***
Random	t=-21.3 df=669.4 p<0.001 ***	t=-16.2 df=644.4 p<0.001 ***	t=-15.2 df=638.5 p<0.001 ***	t=-13.5 df=620.3 p<0.001 ***	–

Table A.1: The full t-test results found when comparing the mean scores for the different methods of creating colour schemes from the "Compare the Results" experiment.

Appendix B

How professional and How artistic were the schemes from each method?

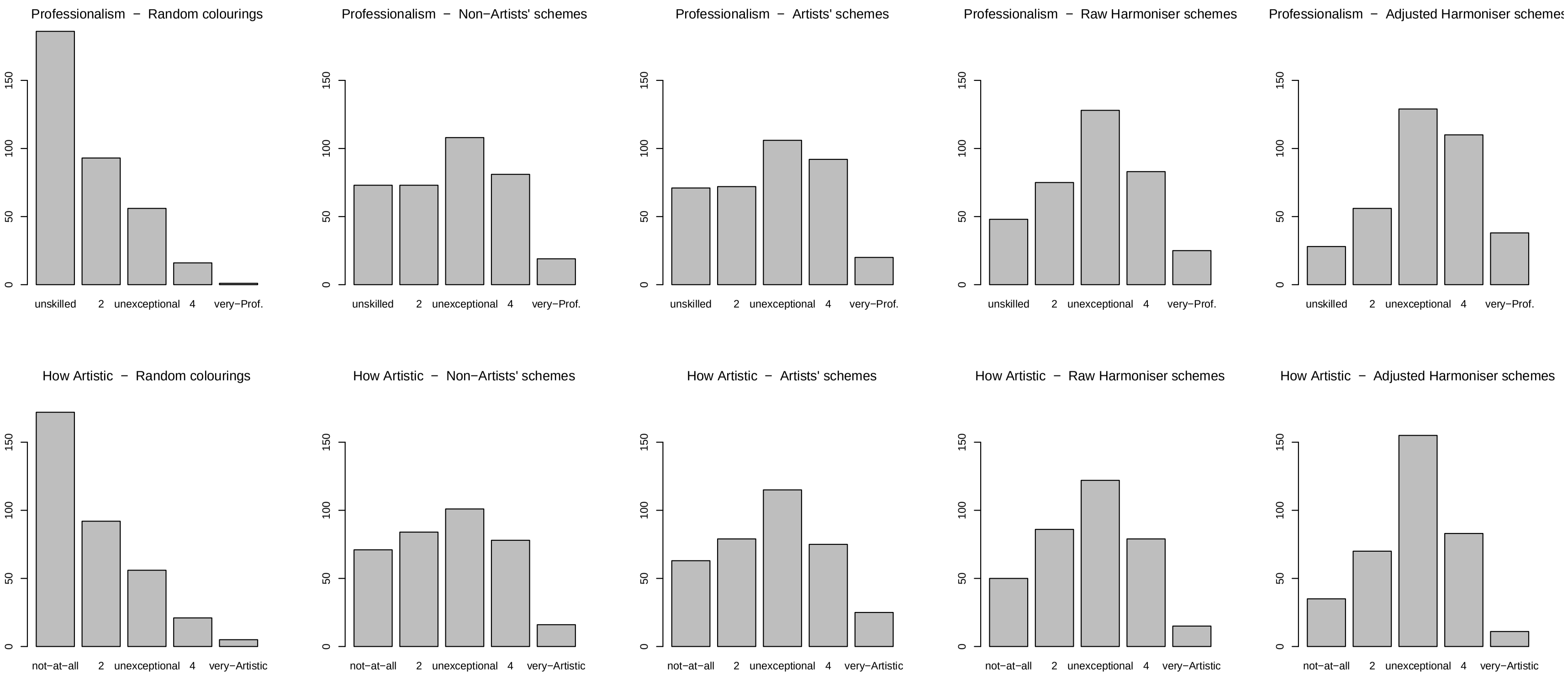


Figure B.1: The participants' evaluation of how-professional (upper) and how-artistic (lower) the colour schemes were for each creation method (from left to right): randomly, by a human non-artist, human artist, Harmoniser-unadjusted, and Harmoniser-adjusted.

Appendix C

Colour-related publications and presentations by the author

The author has published and presented work related to colour prior to the publication of this thesis. These include:

Conference presentations and publications

- Lyons, P., Moretti, G., & Wilson, M. (2000). *Colour Group Selection for Computer Interfaces*. Proc. Human Vision and Electronic Imaging V (Proceedings of SPIE Volume 3959), San Jose, USA. 302–313.
- Moretti, G., & Lyons, P. (2000). *Harmonious Colour Selection for User Interfaces*. Symposium on Computer-Human Interaction, ACM SIGCHI (NZ) - Association of Computing Machinery, Special Interest Group on Human Computer Interaction, Palmerston North, New Zealand.
- Moretti, G.; Lyons, P. & Wilson, M. (2000). *Chromatic Interpolation for Interface Design*. OZCHI 2000: Interfacing reality in the new millennium. Annual Conference of CHISIG, the Computer Human Interaction Special Interest Group of the Ergonomics Society of Australia, Sydney, Australia. 154–161.
- Lyons, P., & Moretti, G. (2001). *Colour Group Selection for Computer Interfaces*. Proc. IHM-HCI 2001 (vol 2). 15th Annual Conference of the Human-Computer Interaction Group of the British Computer Society - Joint AFIHM-BCS Conference on Human Computer Interaction, Lille, France. 21–24.
- Moretti, G., & Lyons, P. (2002). *Tools for the Selection of Colour Palettes*. Proc. SIGCHI-NZ Symposium on Computer-Human Interaction, Hamilton, New Zealand. 11–12.
- Lyons, P., & Moretti, G. (2004). *Nine Tools for Generating Harmonious Colour Schemes*. Proc. 6th Asia Pacific Conference on Computer Human Interaction, APCHI 2004. Rotorua, New Zealand. 241–251.
- Moretti, G., Lyons, P., & Wilson, M. (2004). *Chromotome: A 3D Interface for Exploring Colour Space*. Proc. 6th Asia Pacific Conference on Computer Human Interaction, APCHI 2004. Rotorua, New Zealand. 283–292.

- Moretti, G., & Lyons, P. (2005). *Controlling the Complexity of Grouped Items in Colour Interfaces.* Proc. 6th International Conference. New Zealand Chapter of the ACM's Special Interest Group on Computer-Human Interaction. (CHINZ 2005): Making CHI Natural, Auckland, New Zealand. 19–24.
- Lyons, P. & Moretti, G. (2005). *Incorporating Groups into a Mathematical Model of Color Harmony for Generating Color Schemes for Computer Interfaces.* Proc. IEEE International Conference on Virtual Environments, Human-Computer Interfaces, and Measurement Systems (VECIMS 2005). 80–85.

Conference tutorials on colour design

- Lyons, P., & Moretti, G. (2000). *The Use of Colour in Interface Design: Tutorial.* OZCHI 2000: Interfacing reality in the new millennium. Annual Conference of CHISIG, the Computer Human Interaction Special Interest Group of the Ergonomics Society of Australia, Sydney, Australia.
- Lyons, P., & Moretti, G. (2001). *Colour in Computer Interfaces.* Proc. IHM-HCI 2001 (vol 2). 15th Annual Conference of the Human-Computer Interaction Group of the British Computer Society - Joint AFIHM-BCS Conference on Human Computer Interaction, Lille, France. 227-228.
- Lyons, P., & Moretti, G. (2001). *Colour in Computer Interfaces.* HCC'01 - IEEE Symposia on Human-Centric Computing Languages and Environments. September 5-7, 2001, Stresa, Italy.
- Lyons, P., & Moretti, G. (2004). *Computers, Applications and Colour.* Separately published under the auspices of 6th Asia Pacific Conference on Computer Human Interaction, APCHI 2004. Rotorua, New Zealand.

Appendix D

A Calculation of Colours – Vincent van Gogh

The name of this thesis is not original. It comes from the following letter that Vincent van Gogh wrote to his brother Theo. From one who stands in awe of what artists can do with colour, it is only appropriate to acknowledge the source. The line "So beautiful is that background that it arose spontaneously from a calculation of colours. Am I wrong in this?" is towards the end of the letter, on page 266.

The painting, "Still Life with Bible" by Vincent Van Gogh, referred to in the letter.

Letter from Vincent van Gogh to Theo van Gogh
Nuenen, late October 1885

Dear Theo,

I read your letter about black with great pleasure, and it convinces me that you have no prejudice against black.

Your description of Manet's study, "The Dead Toreador", was well analyzed. And the whole letter proves the same as your sketch of Paris suggested to me at the time, that if you put yourself to it, you can paint a thing in words.

It is a fact that by studying the laws of the colours, one can move from instinctive belief in the great masters to the analysis of why one admires – what one admires – and that indeed is necessary nowadays, when one realizes how terribly arbitrary and superficially people criticize.

You must just let me maintain my pessimism about the art trade as it is these days, for it does not at all include discouragement. This is my way of reasoning. Supposing

I am right in considering that curious haggling about prices of pictures to be more and more like the bulb trade. I repeat, supposing that like the bulb trade at the end of the last century, so the art trade, along with other branches of speculation at the end of this century, will disappear as they came, namely rather quickly. The bulb trade may disappear – the flower-growing remains. And I for myself am contented, for better or for worse, to be a small gardener, who loves his plants.

Just now my palette is thawing and the frigidness of the first beginning has disappeared.

It is true, I often blunder still when I undertake a thing, but the colours follow of their own accord, and taking one colour as a starting-point, I have clearly before my mind what must follow, and how to get life into it.

Jules Dupré is in landscape rather like Delacroix, for what enormous variety of mood did he express in symphonies of colour.

Now a marine, with the most delicate blue-greens and broken blue and all kinds of pearly tones, then again an autumn landscape, with a foliage from deep wine-red to vivid green, from bright orange to dark havana, with other colours again in the sky, in greys, lilacs, blues, whites, forming a further contrast with the yellow leaves.

Then again a sunset in black, in violet, in fiery red.

Then again, more fantastic, what I once saw, a corner of a garden by him, which I have never forgotten: black in the shadow, white in the sun, vivid green, a fiery red and then again a dark blue, a bituminous greenish brown, and a light brown-yellow. Colours that indeed have something to say for themselves.

I have always been very fond of Jules Dupré, and he will become still more appreciated than he is. For he is a real colourist, always interesting, and so powerful and dramatic.

Yes, he is indeed a brother to Delacroix.

As I told you, I think your letter about black very good, and what you say about not painting local colour is also quite correct. But it doesn't satisfy me. In my opinion there is much more behind that not painting local colour.

"Les vrais peintres sont ceux qui ne font pas la couleur locale" [The true painters are those who do not render local colour] – that was what Blanc and Delacroix discussed once.

May I not boldly take it to mean that a painter does better to start from the colours on his palette than from the colours in nature? I mean, when one wants to paint, for instance, a head, and sharply observes the reality one has before one, then one may think: that head is a harmony of red-brown, violet, yellow, all of them broken – I will put a violet and a yellow and a red-brown on my palette and these will break each other.

I retain from nature a certain sequence and a certain correctness in placing the tones, I study nature, so as not to do foolish things, to remain reasonable – however, I don't mind so much whether my colour corresponds exactly, as long as it looks beautiful on my canvas, as beautiful as it looks in nature.

Far more true is a portrait by Courbet, manly, free, painted in all kinds of beautiful deep tones of red-brown, of gold, of colder violet in the shadow with black as repoussoir, with a little bit of tinted white linen as a repose to the eye – finer than a portrait by whomever you like, who has imitated the colour of the face with horribly close precision.

A man's head or a woman's head, well contemplated and at leisure, is divinely beautiful, isn't it? Well, that general harmony of tones in nature, one loses it by painfully exact imitation, one keeps it by recreating in an equivalent colour range, that may be not exactly or far from exactly like the model.

Always and intelligently to make use of the beautiful tones which the colours form of their own accord, when one breaks them on the palette, I repeat – to start from one's palette, from one's knowledge of colour-harmony, is quite different from following nature mechanically and obsequiously.

Here is another example: suppose I have to paint an autumn landscape, trees with yellow leaves. All right – when I conceive it as a symphony in yellow, what does it matter whether the fundamental colour of yellow is the same as that of the leaves or not? It matters very little.

Much, everything depends on my perception of the infinite variety of tones of one and the same family.

Do you call this a dangerous inclination towards romanticism, an infidelity to "realism,"a "peindre de chic", [painting without copying reality] a caring more for the palette of the colourist than for nature? Well, que soit. Delacroix, Millet, Corot, Dupré, Daubigny, Breton, thirty names more, are they not the heart and soul of the art of painting of this century, and are they not all rooted in romanticism, though they surpassed romanticism?

Romance and romanticism are of our time, and painters must have imagination and sentiment. Luckily realism and naturalism are not free from it. Zola creates, but does not hold up a mirror to things, he creates wonderfully, but creates, poetises, that is why it is so beautiful. So much for naturalism and realism, which nonetheless stand in connection to romanticism.

And I repeat that I am touched when I see a picture of about the years '30-'48, a Paul Huet, an old Israëls, like the "Fisherman of Zandvoort", a Cabat, an Isabey.

But I find so much truth in that saying: "ne pas peindre le ton local", [do not paint local tone] that I far prefer a picture in a lower tonal scale than nature to one which is exactly like nature.

Rather a watercolour that is somewhat vague and unfinished than one which is worked up to simulate reality.

That saying: "ne pas peindre le ton local", has a broad meaning, and it leaves the painter free to seek for colours which form a whole and harmonize, which stand out the more in contrast to another combination.

What do I care whether the portrait of an honourable citizen tells me exactly the milk-and-watery bluish, insipid colour of that pious man's face – which I would never have noticed. But the citizens of the small town, where the above-mentioned individual has rendered himself so meritorious that he thought himself obliged to impress his physiognomy on posterity, are highly edified by the correct exactness.

Colour expresses something by itself one cannot do without this, one must use it; that which is beautiful, really beautiful – is also correct; when Veronese had painted the portraits of his beau-monde in the “Marriage at Cana”, he had spent on it all the richness of his palette in sombre violets, in splendid golden tones. Then – he thought still of a faint azure and a pearly-white – which does not appear in the foreground. He detonated it on in the background – and it was right, spontaneously it changes into the ambience of marble palaces and sky, which characteristically consummates the ordering of the figures.

So beautiful is that background that it arose spontaneously from a calculation of colours. Am I wrong in this?

Is it not painted differently than somebody would do it who had thought at the same time of the palace and of the figures as one whole?

All that architecture and sky is conventional and subordinate to the figures, it is calculated to make the figures stand out beautifully.

Surely that is real painting, and the result is more beautiful than the exact imitation of the things themselves. To think of one thing and to let the surroundings belong to it and proceed from it.

To study from nature, to wrestle with reality – I don’t want to do away with it, for years and years I myself have been so engaged, almost fruitlessly and with all kinds of sad results.

I should not like to have missed that error.

I mean that it would be foolish and stupid always to go on in that same way, but not that all the pains I took should be absolutely dismissed.

“On commence par tuer, on finit par guerir”, [One begins by killing, one ends by healing] is a doctor’s saying. One starts with a hopeless struggle to follow nature, and everything goes wrong; one ends by calmly creating from one’s palette, and nature agrees with it, and follows. But these two contrasts are not separable from one another. The drudging, though it may seem in vain, gives an intimacy with nature, a sounder knowledge of things. And a beautiful saying by Doré (who sometimes is so clever) is: je me souviens. [I remember.] Though I believe that the best pictures are more or less freely painted by heart, still I cannot divorce the principle that one can never study and toil too much from nature. The greatest, most powerful imaginations have at the same time made things directly from nature which strike one dumb.

In answer to your description of the study by Manet, I send you a still-life of an open – so a broken white – Bible bound in leather, against a black background, with a yellow-brown foreground, with a touch of citron yellow.

I painted that in one rush, during a single day.

This to show you that when I say that I have perhaps not plodded completely in vain, I dare say this, because at present it comes quite easily to me to paint a given subject unhesitatingly, whatever its form or colour may be. Recently I painted a few studies out of doors, of the autumn landscape. I shall write again soon, and send this letter in haste to tell you that I was quite pleased with what you say about black.

Goodbye,
Yours, Vincent

At this time, Vincent was 32 years old.

Source: Vincent van Gogh. Letter to Theo van Gogh. Written late October 1885 in Nuenen. Translated by Mrs. Johanna van Gogh-Bonger, edited by Robert Harrison, published in *The Complete Letters of Vincent van Gogh*, Publisher: Bulfinch, 1991, number 429. `http://webexhibits.org/vangogh/letter/15/429.htm`.

Text of van Gogh's letter used with permission from the "van Gogh's Letters" exhibit at `www.WebExhibits.org` by `http://www.IDEA.org` (URLs accessed May 21, 2010).

Bibliography

Abramov, I. and Gordon, J. (**1994**). "Color Appearance: on seeing red – or yellow, or green, or blue." *Annual Review of Psychology*, 45: 451–485.

Adobe Systems (**2005**). "Adobe RGB (1998) Color Image Encoding." URL `http://www.adobe.com/digitalimag/pdfs/AdobeRGB1998.pdf`. Accessed May 20, 2010.

Adobe Systems (**2009**). "Kuler." URL `http://kuler.adobe.com`. Accessed May 19, 2010.

Albers, J. (**1963**). *The Interaction of Color*. Yale University Press, New Haven CT.

Albers, J. (**1969**). *Search versus Re-Search*. Hartford.

Alexander, K. R. and Shansky, M. S. (**1976**). "Influence Of Hue, Value, And Chroma On Perceived Heaviness Of Colors." *Perception & Psychophysics*, 19(1): 72–74.

Alleysson, D. and Hérault, J. (**1997**). "Photoreceptor nonlinearities can account for the MacAdam ellipses." In "Proc. European Conference of Visual Perception," volume 26.

Amato, M. and Inaebnit, D. (**1991**). "Clinical usefulness of high intensity green light phototherapy in the treatment of neonatal jaundice." *European Journal of Pediatrics*, 150(4): 274–276.

Anderson, M., Motta, R., Chandrasekarand, S., and Stokes, M. (**1995**). "Proposal for a standard default color space for the InternetsRGB." In "Proceedings of the 4th Joint Conference on Color Imaging Color, Science, Systems and Applications," pages 238–245. Society for Imaging Science & Technology (IS&T).

Aristotle (**2010**). "Meterorologica (Meteorology)." URL `http://classics.mit.edu/Aristotle/meteorology.html`. Accessed May 21, 2010. Translated by E. W. Webster. Original 350BC.

Arnheim, R. (**1987**). "Progress in Color Composition." *Leonardo*, 20(2): 165–168. URL `http://www.jstor.org/stable/1578333`. Accessed May 21, 2010.

Bailey, P., Manavazhi, M., and Olomolaiye, P. (**2001**). "A Conceptual Framework for a Decision Support System for Colour Selection and Use in the Design of the Built Environment." In "IV '01: Proceedings of the Fifth International Conference on Information Visualisation," page 213. IEEE Computer Society, Washington, DC, USA. URL `http://doi.ieeecomputersociety.org/10.1109/IV.2001.942061`. Accessed May 20, 2010.

Bailey, P., Manktelow, K., and Olomolaiye, P. (**2003**). "Examination of the color selection process within digital design for the built environment." In "Theory and Practice of Computer Graphics, 2003. Proceedings," pages 193–200.

Ball, P. (**2001**). *Bright Earth: The Invention of Colour.* Penguin Books, London.

Bangor, A. W. (**1998**). *Improving Access to Computer Displays: Readability for Visually Impaired Users.* Master's thesis, Virginia Polytechnic Institute and State University.

Barker, L. (**1999**). *Changing Rooms: Colour.* BBC Worldwide.

Barnes, C. S. and Shevell, S. K. (**2002**). "Simultaneous S-cone Contrast." *Vision Research*, 42(1): 75–88.

Barnes, C. S., Wei, J. P., and Shevell, S. K. (**1999**). "Chromatic induction with remote chromatic contrast varied in magnitude, spatial frequency, and chromaticity." *Vision Research*, 39(21): 3561–3574.

Bauersfeld, F., Penny and Slater, L., Jodi (**1991**). "User-oriented color interface design: direct manipulation of color in context." In "CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems," pages 417–418. ACM Press, New York, NY, USA. URL `http://doi.acm.org/10.1145/108844.108981`. Accessed May 20, 2010.

Beckes-Malorny, U. (**2007**). *Wassily Kandinsky, 1866–1944, the journey to abstraction.* Taschen.

Bentley, P. J. and Corne, D. W. (**2002**). *Creative Evolutionary Systems.* Academic Press.

Beretta, G. (**1994**). "Functional Color Selection System." US Patent 5311212.

Beretta, G. B. (**1993**). "Reference Color Selection System." US Patent 5254978.

Birren, F. (**1961**). *Color, Form and Space.* Reinhold Publishing Corporation.

Birren, F. (**1969**a). *Munsell: A Grammar of Color.* Van Nostrand-Reinhold.

Birren, F. (**1969**b). *Principles of Color: A review of past traditions and modern theories of color harmony.* Van Nostrand Reinhold.

Bloch, P. H. (**1995**). "Seeking the Ideal Form: Product Design and Consumer Response." *Journal of Marketing*, 59(3): 16–29. URL `http://links.jstor.org/sici?sici=0022-2429%28199507%2959%3A3%3C16%3ASTIFPD%3E2.0.CO%3B2-3`. Accessed May 21, 2010.

Boyatzis, C. and Varghese, R. (**1994**). "Childrens emotional associations with colors." *Journal of Genetic Psychology*, 155(1): 77–89.

Brenner, E., Ruiz, J. S., Herraiz, E. M., Cornelissen, F. W., and Smeets, J. B. J. (**2003**). "Chromatic induction and the layout of colours within a complex scene." *Vision Research*, 43(13): 1413–1421.

Buchner, A. and Baumgartner, N. (**2007**). "Text-background polarity affects performance irrespective of ambient illumination and colour contrast." *Ergonomics*, 50(7): 1036–1063. URL `http://www.informaworld.com/10.1080/00140130701306413`. Accessed May 20, 2010.

Burchett, K. E. (**1991**). "Color harmony attributes." *Color Research & Application*, 16(4): 275–278.

Burchett, K. E. (**2002**). "Color harmony." *Color Research & Application*, 27(1): 28–31. URL `http://dx.doi.org/10.1002/col.10004`. Accessed May 20, 2010.

Caivano, J. L. and López, M. A. (**2003**). "The Rhetoric of Black, White and Red: reasonability and aesthetics to persuade with color." In "Proceedings of AIC 2003 – "Color Communication and Management"," pages 341–346. URL `http://www.fadu.uba.ar/sitios/sicyt/color/2003AICr.pdf`. Accessed May 20, 2010.

Camgöez, N., Yener, C., and Güven, D. (**2002**). "Effects of Hue, Saturation, and Brightness on Preference." *Color Research and Application*, 27(3): 199–207.

Campadelli, P., Schettini, R., and Zuffi, S. (**2001**). "A System for the Automatic Selection of Conspicuous Color Sets for Qualitative Data Display." *IEEE Transactions on Geoscience and Remote Sensing*, 39(10): 2283–2286.

Chevreul (**1855**). *The Principles of Harmony and Contrast of Colours and their Application to the Arts*. London Longman, Brown, Green and Longmans.

Chuang, M.-C. and Ma, Y.-C. (**2001**). "Expressing the expected product images in product design of micro-electronic products." *International Journal of Industrial Ergonomics*, 27(4): 233–245. URL `http://dx.doi.org/10.1016/S0169-8141(00)00053-6`. Accessed May 20, 2010.

CIE (**1926**). *Commission Internationale de l'Eclairage proceedings, 1924*. Cambridge Unversity Press.

Cooper, E. and Kamei, K. (**1999**). "Color Balance in Color Placement Support Systems for Visualization." In "Proceedings of HCI International (the 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Ergonomics and User Interfaces-Volume I," pages 486–490. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA.

Cooper, E. and Kamei, K. (**2001**). "Implementing the color conspicuity model for visualization." In "Joint 9th IFSA World Congress and 20th NAFIPS International Conference," volume 3, pages 1450–1453. URL `http://dx.doi.org/10.1109/NAFIPS.2001.943762`. Accessed May 20, 2010.

Cooper, E. W. and Kamei, K. (**2002**). "A study of color conspicuity for ease-of-use inference in visualization." *Color Research & Application*, 27(2): 74–82. URL `http://dx.doi.org/10.1002/col.10033`. Accessed May 20, 2010.

Coulson, K. (**1998**). *Graphics: A New Zealand Approach*. Addison Wesley Longman.

Damhorst, M. L. and Reed, J. A. P. (**1986**). "Clothing Color Value and Facial Expression - Effects on Evaluations Of Female Job Applicants." *Social Behavior & Personality: An International Journal*, 14(1): 89 – 98.

D'Andrade, R. and Egan, M. (**1974**). "The Colors of Emotion." *American Ethnologist*, 1(1): 49–63. URL `http://www.jstor.org/stable/643802`. Accessed May 20, 2010.

Dorr, S. and Neumeyer, C. (**1997**). "Simultaneous color contrast in goldfish - A quantitative study." *Vision Research*, 37(12): 1581–1593.

Douglas, S. and Kirkpatrick, T. (**1996**). "Do Color Models Really Make a Difference?" In "Proceedings of the SIGCHI conference on Human factors in computing systems," pages 399–405. ACM Press. URL `http://portal.acm.org/citation.cfm?id=238585`. Accessed May 21, 2010.

Douglas, S. A. and Kirkpatrick, A. E. (**1999**). "Model and Representation: The Effect of Visual Feedback on Human Performance in a Color Picker Interface." *ACM Transactions on Graphics*, 18(2): 96–127.

Eckert, C., Kelly, I., and Stacey, M. (**1999**). "Interactive generative systems for conceptual design: An empirical perspective." *Artificial Intelligence for Engineering Design Analysis and Manufacturing*, 13(4): 303–320.

Edwards, B. (**1993**). *Drawing on the Right Side of the Brain*. Harper Collins, revised edition.

Ennever, J. (**1990**). "Blue light, green light, white light, more light: treatment of neonatal jaundice." *Clinics in perinatology*, 17(2): 467–481.

Evans, R. (**1943**). "Visual Processes and Color Photography." *Journal of the Optical Society of America*, 33: 579–614.

Fairchild, M. D. (**1998**). *Color Appearance Models*. Prentice-Hall.

Felton, M. J. (**2000**). "Survival of the Fittest in Drug Design." *Modern Drug Discovery*, 3(9). URL `http://pubs.acs.org/subscribe/journals/mdd/v03/i09/html/felton.html`. Accessed May 20, 2010.

Fogel, D. B. (**2005**). *Evolutionary computation: toward a new philosophy of machine intelligence*. Wiley-IEEE Press, 3rd edition.

Foster, V., editor (**2004**). *Colour Matching Handbook*. Grange Books.

Fukuzumi, S. and Hayashi, Y. (**1989**). "Luminance and stimulus purity of VDT display color in terms of readability." *International Journal of Human-Computer Interaction*, 1(1): 115–135.

Fukuzumi, S., Yamazaki, S., Kamijo, T., and Hayashi, Y. (**1998**). "Physiological and psychological evaluation for visual display colour readability: A visual evoked potential study and a subjective evaluation study." *Ergonomics*, 41(1): 89–108. URL `http://www.informaworld.com/index/OJHCYP6GAUYYWP74.pdf`. Accessed May 21, 2010.

Gage, J. (**2000**). *Colour and Meaning: Art, Science and Symbolism*. Thames and Hudson, London.

Gamma, E., Helm, R., Johnson, R., and Vlissides, J. (**1994**). *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional.

Garcia, M. L. and Caldera, C. I. (**1996**). "The effect of color and typeface on the readability of on-line text." *Computers & Industrial Engineering*, 31(1-2): 519–524. URL `http://dx.doi.org/10.1016/0360-8352(96)00189-1`. Accessed May 20, 2010.

Gavel, J. (**1979**). *Colour - A study of its Position in the Art Theory of the Quattro & Cinquecento*. Rosenlundstryckeriet AB, Stockholm.

Giarrantano, J. and Riley, G. (**1998**). *Expert Systems: principles and programming*. PWS Publishing, 3rd edition.

Gimbel, T. (**1993**). *The Colour Therapy Workbook*. Element Books.

Goldstein, H. and Goldstein, V. (**1961**). *Art in Everyday Life*. The Macmillan Company, 6th edition.

Gradisar, M., Humar, I., and Turk, T. (**2007**). "The Legibility of Colored Web Page Texts." In "29th International Conference on Information Technology Interfaces, 2007. ITI 2007.", pages 233–238. URL TheLegibilityofColoredWebPageTexts. Checked May 21, 2010.

Granger, G. (**1955**a). "Aesthetic Measure Applied to Color Harmony: An Experimental Test." *The Journal of General Psychology*, 52: 205–212.

Granger, G. (**1955**b). "An experimental study of color preferences." *Journal of General Psychology*, 52: 3–20.

Granger, G. W. (**1953**). "Area balance in color harmony; an experimental study." *Science*, 117(3029): 59–61.

Graves, M. (**1952**). *Color Fundamentals - with 100 color schemes.* McGraw Hill.

Grimsaeth, K. (**2007**). "Kansei Engineering - Linking emotions and product features." Technical report, Norwegian University of Science and Technology (NTNU).

Guan, S.-S. (**2002**). "A Study of Color Harmony relating with Color Area." In R. Chung and A. Rodrigues, editors, "Proc. SPIE - 9th Congress of the International Colour Association," volume 4421, pages 199–202.

Guild, J. (**1931**). "The Colorimetric Properties of the Spectrum." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 230: 149–187. URL http://www.jstor.org/stable/91229. Accessed May 21, 2010.

Gulliksen, J. and Sandblad, B. (**1995**). "Domain-Specific Design of User Interfaces." *International Journal of Human-Computer Interaction*, 7(2): 135–151.

Gurura, H., MacDonald, L., and Dalke, H. (**2004**). "Background: An essential factor in colour harmony." In "AIC 2004 Color and Paints, Interim Meeting of the International Color Association," .

Hall, R. H. and Hanna, P. (**2004**). "The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention." *Behaviour & Information Technology*, 23(3): 183 – 195.

Hård, A. and Sivik, L. (**2001**). "A theory of colors in combination: a descriptive model related to the NCS color-order system." *Color Research & Application*, 26(1): 4–28. URL http://dx.doi.org/10.1002/1520-6378(200102)26:1<4::AID-COL3>3.0.CO;2-T. Accessed May 20, 2010.

Healey, C. G. (**1996**). "Choosing effective colours for data visualization." In "VIS '96: Proceedings of the 7th conference on Visualization '96," pages 263–ff. IEEE Computer Society Press, Los Alamitos, CA, USA.

Hering, E. (**1964**). *Outlines of a theory of light sense.* Harvard University Press. Translated by Leo M. Hurvich and Dorothea Jameson. Original work published in 1874.

Hilscher, M. (**2005**). *Performance Implications of Alternative Color-Codes in Airport X-Ray Baggage Screening.* Ph.D. thesis, The University of Central Florida.

Holland, J. (**1975**). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

Holland, J. (**2009**). "Genetic Algorithms." URL `http://www.econ.iastate.edu/tesfatsi/holland.GAIntro.htm`. Accessed May 20, 2010.

Hornung, D. (**2005**). *Colour - a workbook for artists and designers*. Laurence King Publishing.

Howlett, V. (**1996**). *Visual Interface Design for Windows*. Wiley Computer Publishing.

Hubel, P. (**1999**). "The Perception of Color at Dusk and Dawn." In "Proceedings of The Seventh Color Imaging Conference: Color Science, Systems, and Applications," pages 48–51.

Hunter Associates Labs (**2008**). "Equivalent White Sources and CIE Illuminants." URL `http://www.hunterlab.com/appnotes/an05%5F05.pdf`. Accessed May 20, 2010.

Hurvich, L. and Jameson, D. (**1957**). "An opponent-process theory of color vision." *Psychological Review*, 64(6 (part 1)): 384–404. URL `http://dx.doi.org/10.1037/h0041403`. Accessed May 20, 2010.

Ichikawa, M., Tanaka, K., Kondo, S., Hiroshima, K., Ichikawa, K., Tanabe, S., and Fukami, K. (**2004**). "Preliminary study on color modification for still images to realize barrier-free color vision." In "Systems, Man and Cybernetics, 2004 IEEE International Conference on," volume 1, pages 36–41.

Ihaka, R. (**2002**). "Colour for Presentation Graphics." In K. Hornik and F. Leisch, editors, "Proceedings of the 3rd Interational Workshop on Distributed Statistical Computing (DSC2003)," .

International Color Consortium (**2004**). "Specification ICC.1:2004-10 (Profile version 4.2.0.0): Image technology colour management." http://www.color.org/ICC1v42%5F2006-05.pdf. URL `http://www.color.org/icc%5Fspecs2.xalter`.

Ishizaki, S. (**1995**). "Color adaptive graphics: what you see in your color palette isn't what you get!" In "CHI '95: Conference companion on Human factors in computing systems," pages 300–301. ACM Press, New York, NY, USA. URL `http://doi.acm.org/10.1145/223355.223684`. Accessed May 20, 2010.

Ishizaki, S. (**1997**). "Adjusting Simultaneous Contrast Effect for Dynamic Information Display." In "Recent Progress in Color Science," pages 68–71. Society of Imaging Science & Technology.

Itten (**1970**a). *The Elements of Color, translated by Enrnst van Hagen, edited by Faber Birren*. Van Nostrand-Reinhold, New York.

Itten, J. (**1970**b). *The Art of Color*. John Wiley and Sons.

Jackson, C. (**1985**). *Colour Me Beautiful*. Little Hills Press.

Jacobson, N. and Bender, W. (**1996**). "Color as a determined communication." *IBM Systems Journal*, 35(3-4): 526–538.

Jameson, D. and Hurvich, L. M. (**1989**). "Essay Concerning Color Constancy." *Annual Review of Psychology*, 40: 1–22.

Jefferson, L. and Harvey, R. (**2006**). "Accommodating color blind computer users." In "Assets '06: Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility," pages 40–47. ACM Press, New York, NY, USA. URL `http://doi.acm.org/10.1145/1168987.1168996`. Accessed May 20, 2010.

Jordan, P. W. (**1998**). "Human factors for pleasure in product use." *Applied Ergonomics*, 29(1): 25–33. URL `http://dx.doi.org/10.1016/S0003-6870(97)00022-7`. Accessed May 20, 2010.

Judd, D. and Wyszecki, G. (**1963**). *Color in Business, Science and Industry.* Wiley, 2nd edition.

Judd, D. B. (**1959**). *Color in Business, Science and Industry.* John Wiley & Sons (New York), Chapman and Hall (London).

Kagawa, T., Nishino, H., and Utsumiya, K. (**2003**). "A Color Design Assistant based on User's Sensitivity." In "Proceedings of IEEE International Conference on Systems, Man and Cybernetics," volume 1, pages 974–979.

Kaiser, P. (**1984**). "Phototherapy using Chromatic, White, and Ultraviolet-Light." *Color Research & Application*, 9(4): 195–205.

Kaiser, P. K. and Boynton, R. M. (**1996**). *Human Color Vision.* Optical Society of America, 2nd edition.

Karvonen, K. (**2000**). "The beauty of simplicity." In "CUU '00: Proceedings on the 2000 conference on Universal Usability," pages 85–90. ACM, New York, USA. URL `http://doi.acm.org/10.1145/355460.355478`. Accessed May 20, 2010.

Katz, B. F. (**1999**). "Color Contrast and Color Preference." *Empirical Studies of the Arts*, 17(1): 1–24.

Kaya, N. and Epps, H. (**2004**). "Relationship between color and emotion: A study of college students." *College Student Journal*, 38(3): 396–418.

Kelly, I. (**1996**). "A Two Stage Evolutionary Model for the Computer Aided Design of Colour Combinations." In L. Sivik, editor, "AIC (Association Internationale de la Couleur) Interim Meeting 96 – Colour and psychology," Scandinavian Colour Institute.

Kelly, I. (**1999**). *An Evolutionary Interaction Approach to Computer-aided Colour Design.* PhD thesis, The Open University, Milton Keynes, UK.

Kim, H.-S. and Cho, S.-B. (**2000**). "Application of Interactive Genetic Algorithm to Fashion Design." *Engineering Applications of Artificial Intelligence*, 13(6): 635–644. URL `http://dx.doi.org/10.1016/S0952-1976(00)00045-2`. Accessed May 20, 2010.

Klassen, R. V., Braun, K., Buckley, R. R., and Janamanchi, K. (**1998**). "Visibility of Thin Lines on Coloured Backgrounds." In "IS&T/SID Sixth Color Imaging Conference: Color Science, Systems and Applications," pages 27–32.

Kobayashi, S. (**1981**). "The Aim and Method of the Color Image Scale." *Color Research and Application*, 6(2): 93–107.

Kobayashi, S. (**1987**). *A Book of Colors.* Kodansha International, New York.

Kobayashi, S. (**1998**). *Colorist.* Kodansha International, New York.

Kondratova, I. and Goldfarb, I. (**2006**). "Cultural Interface Design: Global Colors Study." In "Lecture Notes in Computer Science: On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops," pages 926–934. Springer Berlin / Heidelberg. URL http://www.springerlink.com/content/352687007t2v0058. Accessed May 20, 2010.

Kondratova, I., Goldfarb, I., Gervais, R., and Fournier, L. (**2005**). "Culturally Appropriate Web Interface Design: Web Crawler Study." In "Proceedings of Computers and Advanced Technology in Education," .

Krause, J. (**2002**). *Color Index: Over 1100 Color Combinations, CMYK and RGB Formulas, for Print and Web Media.* HOW Design Books (F&W Publications).

Kuehni, R. G. (**2001**). "Uniform Color Space is Not Homogeneous." http://www.iscc.org/aic2001/abstracts/oral/Kuehni.doc. URL http://www.iscc.org/aic2001/abstracts/oral/Kuehni.doc. Accessed May 21, 2010.

Kuehni, R. G. (**2003**). *Color Space and its Divisions.* John Wiley & Sons, Hoboken, New Jersey.

Kuehni, R. G. (**2004**). "Variability in Unique Hue Selection: A Surprising Phenomenon." *Color Research & Application*, 29(2): 158–162. URL http://dx.doi.org/10.1002/col.10237. Accessed May 21, 2010.

Kurosu, M. and Kashimura, K. (**1995**). "Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability." In "CHI '95: Conference companion on Human factors in computing systems," pages 292–293. ACM, New York, USA. URL http://doi.acm.org/10.1145/223355.223680. Accessed May 20, 2010.

Laugwitz, B. (**2001**). "A Web Experiment on Colour Harmony Principles Applied to Computer User Interface Design." In U. D. Reips and M. Bosnjak, editors, "Dimensions of Internet Science," Pabst Science Publishers, Lengerich, Germany.

Leichsenring, F. (**2004**). "The Influence of Color on Emotions in the Holtzman Inkblot Technique." *European Journal of Psychological Assessment*, 20(2): 116–123.

Lin, C.-C. (**2003**). "Effects of contrast ratio and text color on visual performance with TFT-LCD." *International Journal of Industrial Ergonomics*, 31(2): 65–72. URL http://dx.doi.org/10.1016/S0169-8141(02)00175-0. Accessed May 20, 2010.

Lin, C.-C. (**2005**). "Effects of screen luminance combination and text color on visual performance with TFT-LCD." *International Journal of Industrial Ergonomics*, 35(3): 229–235. URL http://dx.doi.org/10.1016/j.ergon.2004.09.002. Accessed May 20, 2010.

Linden, D. S. (**2002**). "Innovative Antenna Design Using Genetic Algorithms." In P. J. Bentley and D. W. Corne, editors, "Creative Evolutionary Systems," pages 487–510. Morgan Kaufmann.

Ling, J. and van Schaik, P. (**2002**). "The effect of text and background colour on visual search of Web pages." *Displays*, 23(5): 223–230. URL http://dx.doi.org/10.1016/S0141-9382(02)00041-0. Accessed May 20, 2010.

Linnett, C. M., Morriss, R. H., Dunlap, W. P., and Fritchie, C. J. (**1991**). "Differences in color balance depending upon mode of comparison." *Journal of General Psychology*, 118(3).

Lleonart, A. (**2009**). *Minimalism and Color DesignSource (Design Source).* Collins Design.

Locher, P., Overbeeke, K., and Stappers, P. J. (**2005**). "Spatial balance of color triads in the abstract art of Piet Mondrian." *Perception*, 34(2): 169–189. URL `http://www.perceptionweb.com/perception/fulltext/p34/p5033.pdf`.

Luke, J. T. and Luke, S. C. (**2000**). "Method for generating numerous harmonious color palettes from two colors." US Patent 6081253.

Lumley, T. (**2006**). "Color coding and color blindness in statistical graphics." *Statistical Computing and Graphics*, 17(2): 4–7.

Luo, M. R. and Gao, X. W. (**1995**). "Quantifying colour appearance. part V. simultaneous contrast." *Color Research & Application*, 20(1): 18–28. URL `http://dx.doi.org/10.1002/col.5080200105`. Accessed May 20, 2010.

Lüscher, M. (**1970**). *The Lüscher colour test (translated and edited by Ian A. Scott – based on the original German text[s] "Psychologie der Farben" by Max Lüscher).* Cape, London.

Lyons, P. and Moretti, G. (**2004**). "Nine Tools for Generating Harmonious Colour Schemes." In M. Masoodian, S. Jones, and B. Rogers, editors, "Proc. 6th Asia Pacific Conference, APCHI 2004: Human Computer Interaction," volume 3101 of *Lecture Notes in Computer Science*, pages 241–251. Springer, Rotorua, New Zealand. URL `http://www.springerlink.com/index/AJCVUT51CQKJCMN8`. Accessed May 20, 2010.

MacAdam, D. (**1942**). "Visual sensitivities to color differences in daylight." *J. Opt. Soc. Am.*, 32: 247–273. URL `http://www.opticsinfobase.org/abstract.cfm?URI=josa-32-5-247`. Accessed May 20, 2010.

MacDonald, L. (**1993**). "Gamut Mapping in Perceptual Colour Space." In "IS & T & SID's Colour Imaging Conference: Transforms and Transportability of Color," pages 193–196.

MacEvoy, B. (**2008**). "J.W. von Goethe's "Zur Farbenlehre"." URL `http://www.handprint.com/HP/WCL/goethe.html`. Accessed May 20, 2010.

MacEvoy, B. (**2009**a). "Color Wheels." URL `http://www.handprint.com/HP/WCL/color13.html`. Accessed May 20, 2010.

MacEvoy, B. (**2009**b). "The Geometry of Color Perception." URL `http://www.handprint.com/HP/WCL/color2.html`. Accessed May 20, 2010.

MacEvoy, B. (**2009**c). "Modern Color Models." URL `http://www.handprint.com/HP/WCL/color7.html`. Accessed May 20, 2010.

MacIntyre, B. (**1991**). "A Constraint-Based Approach to Dynamic Colour Management for Windowing Interfaces." Technical Report CS-91-55, University of Waterloo. URL `http://www.cs.uwaterloo.ca/research/tr/1991/55/tech.pdf`. Accessed May 20, 2010.

Mahyar, F., Cheung, V., Westland, S., and Henry, P. (**2007**). "Investigation of Complementary Colour Harmony in CIELAB Colour Space." In "Proceedings of AIC 2007 – "Color Science for Industry"," pages 82–85.

Manniesing, A. (2003). *Creating Harmonious and Legible Colour Schemes in the Automated Generation of Multimedia Presentations*. MSc thesis - published as a report, Centrum voor Wiskunde en Informatica. URL `http://repository.cwi.nl/search/fullrecord.php?publnr=4088`. Accessed May 20, 2010.

Marcus, A. (1997). "Human Computer Interaction." In M. Helander, T. Landauer, and P. Prabhu, editors, "Handbook of Human-Computer Interaction," chapter Graphical user interface, pages 423–440. Elsevier Science, B.V.

Marsland, S. (2009). *Machine Learning: an algorithmic perspective*. Chapman and Hall/CRC Press.

Martindale, C. (2001). "How does the brain compute aesthetic preference?" *The General Psychologist*, 36(2): 25–35.

McCamy, C. S. (1998). "On the number of discernible colors." *Color Research & Application*, 23(5): 337. URL `http://dx.doi.org/10.1002/(SICI)1520-6378(199810)23:5<337::AID-COL11>3.0.CO;2-6`. Accessed May 20, 2010.

Meier, B., Spalter, A., and Karelitz, D. (2004). "Interactive Color Palette Tools." *Computer Graphics and Applications, IEEE*, 24(3): 64–72.

Meier, B. J. (1988). "ACE: a color expert system for user interface design." *Proceedings of the 1st annual ACM SIGGRAPH symposium on User Interface Software*, pages 117–128. URL `http://doi.acm.org/10.1145/62402.62424`. Accessed May 20, 2010.

Mendoza, H. R. (2004). *Pale Intrusions into Blue: the development of a color*. Master's thesis, The Florida State University, School of Fine Arts and Dance. URL `http://etd.lib.fsu.edu/theses/available/etd-11152004-155644`. Accessed May 20, 2010.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., and Read, G. M. (2001). "Psychological testing and psychological assessment: A review of evidence and issues." *American Psychologist*, 56(2): 128–165.

Meyers-Levy, J. and Peracchio, L. A. (1995). "Understanding the Effects of Color: How the Correspondence between Available and Required Resources Affects Attitudes." *The Journal of Consumer Research*, 22(2): 121–138. URL `http://links.jstor.org/sici?sici=0093-5301%28199509%2922%3A2%3C121%3AUTEOCH%3E2.0.CO%3B2-%23`. Accessed May 20, 2010.

Miyahara, E., Smith, V. C., and Pokorny, J. (1993). "How Surrounds Affect Chromaticity Discrimination." *Journal Of The Optical Society Of America A-Optics Image Science And Vision*, 10(4): 545–553.

Monnier, P. and Shevell, S. K. (2004). "Chromatic induction from S-cone patterns." *Vision Research*, 44(9): 849–856.

Monsef IV, D. A. (2009). "Color Lovers." URL `http://www.colorlovers.com`. Accessed May 20, 2010.

Moon, P. and Spencer, D. (1944a). "Area in Color Harmony." *Journal of Optical Society of America*, 34: 93–103.

Moon, P. and Spencer, D. E. (**1944**b). "Aesthetic measure applied to color harmony." *J. Opt. Soc. Am.*, 34(4): 234–242. URL http://www.opticsinfobase.org/abstract.cfm?URI=josa-34-4-234. Accessed May 20, 2010.

Moon, P. and Spencer, D. E. (**1944**c). "Geometric formulation of classical color harmony." *Journal of the Optical Society of America*, 34(1): 46–59. URL http://dx.doi.org/10.1364/JOSA.34.000046. Accessed May 20, 2010.

Moretti, G. and Lyons, P. (**2002**). "Tools for the Selection of Colour Palettes." In S. Jones and M. Masoodian, editors, "Proc. SIGCHI-NZ Symposium on Computer-Human Interaction," pages 13–18. Hamilton, New Zealand.

Moretti, G. and Lyons, P. (**2005**). "Controlling the Complexity of Grouped Items in Colour Interfaces." In B. Plimmer and R. Amor, editors, "Proc. 6th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction (CHINZ 2005): making CHI natural," pages 19–24. ACM Press, Auckland, New Zealand. URL http://doi.acm.org/10.1145/1073943.1073948. Accessed May 20, 2010.

Moretti, G., Lyons, P., and Wilson, M. (**2004**). "Chromotome: A 3D Interface for Exploring Colour Space." In M. Masoodian, S. Jones, and B. Rogers, editors, "Proc. 6th Asia Pacific Conference, APCHI 2004: Human Computer Interaction," volume 3101 of *Lecture Notes in Computer Science*, pages 283–292. Springer, Rotorua, New Zealand. URL http://www.springerlink.com/index/KL88MNX38LRFNP3J. Accessed May 20, 2010.

Morriss, R. H. and Dunlap, W. P. (**1987**). "Influence of Value on Spatial Balance of Color Pairs." *Journal of General Psychology*, 114(4): 353–361.

Morriss, R. H., Dunlap, W. P., and Hammond, S. E. (**1982**). "Influence of Chroma on Spatial Balance of Complementary Hues." *The American Journal of Psychology*, 95(2): 323–332. URL http://links.jstor.org/sici?sici=0002-9556%28198222%2995%3A2%3C323%3AIOCOSB%3E2.0.CO%3B2-9. Accessed May 21, 2010.

Morton, J. (**2009**a). "Global Color: Clues and Taboos." URL http://www.colorvoodoo.com/cvoodoo2%5Fglobalcolor.html. Accessed May 20, 2010.

Morton, J. and Peterson, L. (**2007**). "Consumer Color Preferences." On Web. URL http://www.colorcom.com/consumercolor.html. Accessed May 21, 2010.

Morton, J. L. (**2009**b). "Colors that Sell: Tried and Tested Color Schemes." URL http://www.colorvoodoo.com/cvoodoo9%5Ftoc.html. Accessed May 20, 2010.

Morton, J. L. (**2009**c). "A Guide to Color Symbolism." URL http://www.colorvoodoo.com/cvoodoo1.html.

Munsell, A. H. (**1905**). *A Color Notation: a measured color system, based on the three qualities Hue, Value and Chroma.* George H. Ellis.

Munsell, A. H. (**1907**). *A Color Notation: a measured color system, based on the three qualities Hue, Value and Chroma.* George H. Ellis, revised edition. URL http://www.gutenberg.org/etext/26054. Accessed May 20, 2010.

Munsell, A. H. (**1912**). "A Pigment Color System and Notation." *The American Journal of Psychology*, 23(2): 236–244. URL http://www.jstor.org/stable/1412843. Accessed May 20, 2010.

Munsell, A. H. and Farnum, R. B. (**1941**). *A Color Notation: An illustrated System Defining All Colors and Their Relations.* Kessinger Publishing.

Murch, G. M. (**1984**). "Physiological principles for the effective use of color." *IEEE Comput. Graph. Appl.*, 4(11): 49–54. URL `http://portal.acm.org/citation.cfm?id=2347.2349`. Accessed May 20, 2010.

Nack, F., Manniesing, A., and Hardman, L. (**2003**). "Colour picking: the pecking order of form and function." In "Proceedings of the eleventh ACM International Conference on Multimedia," pages 279 – 282. ACM. URL `http://doi.acm.org/10.1145/957013.957071`. Accessed May 20, 2010.

Nakakoji, K., Reeves, B. N., Aoki, A., Suzuki, H., and Mizushima, K. (**1995**). "eMMaC: knowledge-based color critiquing support for novice multimedia authors." In "Proceedings of the third ACM international conference on Multimedia," pages 467–476. ACM Press, San Francisco, California, United States. URL `http://doi.acm.org/10.1145/217279.215311`. Accessed May 20, 2010.

Nemcsics, A. (**2009**). "Experimental determination of the laws of color harmony. Part 4: Color preference and the color harmony content." *Color Research & Application*, 34(3): 210–224. URL `http://dx.doi.org/10.1002/col.20489`. Accessed May 20, 2010.

Ngo, D. C. L., Teo, L. S., and Byrne, J. G. (**2002**). "Evaluating interface esthetics." *Knowl. Inf. Syst.*, 4(1): 46–79. URL `http://dx.doi.org/10.1007/s10115-002-8193-6`. Accessed May 20, 2010.

Nielsen, J. and Hahir, M. (**2002**). *Homepage Usability: 50 Website Deconstructed.* New Riders Publishing.

Nielsen, M. and Stokes, M. (**1998**). "The Creation of the sRGB ICC Profile." In "The Sixth Color Imaging Conference: Color Science, Systems, and Applications," page 253.

Noorlander, C. and Koenderink, J. (**1983**). "Spatial and temporal discrimination ellipsoids in color space." *J Opt Soc Am*, 73(11): 1533–43. URL `http://lib.bioinfo.pl/pmid:6644399`.

Norman, R. B. (**1990**). *Electronic Color - the art of color applied to graphic computing.* Van Nostrand Reinhold.

Ojanpaa, H. and Nasanen, R. (**2003**). "Effects of luminance and colour contrast on the search of information on display devices." *Displays*, 24(4-5): 167–178. URL `http://dx.doi.org/10.1016/j.displa.2004.01.003`. Accessed May 20, 2010.

Ostwald, W. (**1969**). *The Color Primer. Edition by Faber Birren from "Die Farbenfibel" (1916).* Von Nostrand Reinhold.

Ou, L.-C. (**2007**). "Existing Theories of Colour Harmony." Web Document - http://colour-emotion.co.uk/harmony.html. URL `http://colour-emotion.co.uk/harmony.html`. Accessed May 20, 2010.

Ou, L.-C., Luo, M. R., and Cui, G. (**2008**). "A Colour Design Tool Based on Empirical Studies." In "Proceedings of DRS2008, Design Research Society Biennial Conference," .

Ou, L.-C., Luo, M. R., Woodcock, A., and Wright, A. (**2004**a). "A Study of Colour Emotion and Colour Preference. Part I: Colour Emotions for Single Colours." *Color Research & Application*, 29(3): 232–240.

Ou, L.-C., Luo, M. R., Woodcock, A., and Wright, A. (**2004**b). "A Study of Colour Emotion and Colour Preference. Part II: Colour Emotions for Two-Colour Combinations." *Color Research & Application*, 29(4): 292–298.

Ou, M. R., Li-Chen amd Luo (**2006**). "A Colour Harmony Model for Two-Colour Combinations." *Color Research and Application*, 31(3): 191–204. URL `http://dx.doi.org/10.1002/col.20208`. Accessed May 20, 2010.

Palffy, Z. (**1976**). "Age-specific characteristics in sensitivity to harmony of colours." *Magyar Pszichologiai Szemle*, 33(6): 562–570.

Paramei, G. V. (**2005**). "Singing the Russian Blues: An Argument for Culturally Basic Color Terms." *Cross-Cultural Research*, 39(1): 10–34. URL `http://ccr.sagepub.com/cgi/content/abstract/39/1/10`. Accessed May 20, 2010.

Pastoureau, M. (**2001**). *Blue : the History of a Color*. Princeton University Press.

Polzella, D. J. and Montgomery, D. A. (**1993**). "Dimensions of Color Harmony." *Bulletin of the Psychonomic Society*, 31(5): 423–425.

Poynton, C. A. (**2003**). *Digital Video and HDTV: Algorithms and Interfaces*. Morgan Kaufmann.

Rapoport, A. (**1994**). "Experimentally elicited judgments of color harmony." *Empirical Studies of the Arts*, 12(1): 63–83.

Rapoport, A. and Rapoport, A. (**1984**a). "Color preferences, color harmony, and the quantitative use of colors." Technical report, University of Haifa, Israel.

Rapoport, A. and Rapoport, A. (**1984**b). "Color preferences, color harmony, and the quantitative use of colors." *Empirical Studies of the Arts*, 2(2): 95–112.

Read, M. A., Sugawara, A. I., and Brandt, J. A. (**1999**). "Impact of space and color in the physical environment on preschool childrens's cooperative behavior." *Environment and Behavior*, 31(3): 413–428.

Ro, Y. M. and Yang, S. (**2004**). "Color adaptation for anomalous trichromats." *International Journal of Imaging Systems and Technology*, 14(1): 16–20. URL `http://dx.doi.org/10.1002/ima.20002`. Accessed May 20, 2010.

Rogowitz, B. E., Kalvin, A. D., Pelah, A., and Cohen, A. (**1999**). "Which Trajectories through which Perceptually Uniform Color Spaces Produce Appropriate Colors Scales for Interval Data?" In "Proc. of Seven Color Imaging Conference: Color Science, Systems, and Applications," pages 321–326.

Rogowitz, B. E. and Rabenhorst, D. A. (**1993**). "CRAFT: A Tool for Customizing Color and Font Selections Guided by Perceptual Rules." In "Proceeding of IS & T Color Imaging Conference: Transforms & Transportability of Color," pages 140–143.

Rogowitz, B. E. and Treinish, L. A. (**1998**). "Data Visualization: the End of the Rainbow." *IEEE Spectrum*, 35(12): 52–59.

Romney, A. K. and Indow, T. (2002). "A model for the simultaneous analysis of reflectance spectra and basis factors of Munsell color samples under D65 illumination in three-dimensional Euclidean space." *PNAS*, 99(17): 11 543–11 546. URL http://www.pnas.org/cgi/content/abstract/99/17/11543. Accessed May 21, 2010.

Ross, B. J., Ralph, W., and Hai, Z. (2006). "Evolutionary Image Synthesis Using a Model of Aesthetics." In "IEEE Congress on Evolutionary Computation," pages 1087–1094. IEEE, Vancouver, Canada.

Runge, P. O. (1810). *Colour-Sphere (Farben-Kugel)*. From Runge, *Die Farben-Kugel*, Hamburg.

Ryall, K. and Marks, J. (1996). "An Interactive System for Drawing Graphs." In "Proceedings of the Symposium on Graph Drawing," pages 387–394. URL http://www.merl.com/publications/TR1996-023/. Accessed May 20, 2010.

Sagawa, K. (1999). "Visual comfort to colored images evaluated by saturation distribution." *Color Research and Application*, 24(5): 313–321.

Salamon, P., Sibani, P., and Frost, R. (2002). *Facts, Conjectures, and Improvements for Simulated Annealing*. Society for Industrial and Applied Mathematics.

Sarapik, V. (1997). "Red: The Colour and the Word." URL http://haldjas.folklore.ee/folklore/vol3/red.htm. Accessed May 21, 2010.

Sawahata, L. E. (1999). *Colour Harmony Workbook: A Workbook and Guide to Creative Colour Combinations*. Rockport Publishers, Massachusetts.

Scandinavian Colour Institute (2009). "NCS Natural Colour System: A Visual Colour System." URL http://www.ncscolour.com. Accessed May 20, 2010.

Scharff, L. F. V. and Ahumada, A. J. (2005). "Understanding text polarity effects." *Perception*, 34: 133–133.

Schauss, A. (1979). "Tranquilizing Effect of Color Reduces Aggressive Behavior and Potential Violence." *Journal of Orthomolecular Psychiatry*, 8(4): 218–221. URL http://www.orthomolecular.org/library/jom/1979/pdf/1979-v08n04-p218.pdf. Accessed May 20, 2010.

Schlueter, K. (1990). *Perceptual Synchronization in Window Systems*. Master's thesis, University of Waterloo.

Schwartz, B. (2005). *The Paradox of Choice: Why More Is Less*. Harper Perennial.

Sharma, G. and Trussell, H. (1997). "Digital color imaging." *IEEE Transactions on Image Processing*, 6(7): 901–932.

Shepherd, A. J. (1999). "Remodelling colour contrast: implications for visual processing and colour representation." *Vision Research*, 39(7): 1329–1345.

Shneiderman, B. (2000). "Creating creativity: user interfaces for supporting innovation." *ACM Trans. Comput.-Hum. Interact.*, 7(1): 114–138. URL http://doi.acm.org/10.1145/344949.345077. Accessed May 20, 2010.

Shneiderman, B., Plaisant, C., Cohen, M., and Jacobs, S. (2009). *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Addison Wesley, 5th edition.

Siegel, S. and Castellan, N. (1988). *Nonparametric Statistics for the Behavioral Sciences.* McGraw Hill.

Silva, S., Madeira, J., and Santos, B. S. (2007). "There is More to Color Scales than Meets the Eye: A Review on the Use of Color in Visualization." In "11th International Conference Information Visualization (IV '07)," volume 00, pages 943–950. IEEE Computer Society, Los Alamitos, CA, USA. URL `http://doi.ieeecomputersociety.org/10.1109/IV.2007.113`. Accessed May 20, 2010.

Singer, J. (1976). *How to Paint in Pastels.* Watson–Guptil Publishers.

Smith, W. J. (1996). *ISO and ANSI Ergonomic Standards for Computer Products: A Guide to Implementation and Compliance.* Prentice-Hall.

Snyder, H. L., Decker, J. J., Lloyd, C. J. C., and Dye, C. (1995). "Effect of Image Polarity on VDT Task Performance." In G. Perlman, G. K. Green, and M. S. Wogalter, editors, "Human Factors Perspectives on Human-Computer Interaction: Selections from Proceedings of Human Factors and Ergonomics Society Annual Meetings, 1983-1994," pages 225–229. Human Factors and Ergomonic Society.

Snyman, J. A. (2005). *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms.* Springer Publishing.

Solano, M. and Joyner, I. (2007). "Performance Analysis of Evolutionary Search with a Dynamic Restart Policy." In "Proceedings of the 20th International Florida Artificial Intelligence Research Society Conference," pages 186–187.

St.Bernard, M. A. (1995). *The effects of color harmony on person perception.* Ph.D. thesis, Michigan State University, USA.

Stockman, A. and Sharpe, L. T. (2000). "Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype." *Vision Research*, 40: 1711–1737. Data available from Colour & Vision Research Laboratory Database, `http://cvision.ucsd.edu`. Accessed May 19, 2010.

Stokes, M., Anderson, M., Chandrasekar, S., and Motta, R. (1996). "A Standard Default Color Space for the Internet — sRGB." URL `http://www.color.org/sRGB.xalter`. Accessed May 20, 2010. Revised – version 1.1 of sRGB proposal.

Stroebel, L., Todd, H., and Zakia, R. (1980). *Visual Concepts for Photographers.* Focal Press.

Sully, J. (1879). "Harmony of Colours." *Mind*, 4(14): 172–191. URL `http://www.jstor.org/stable/2246666`. Accessed May 20, 2010.

Süsstrunk, S., Buckley, R., and Swen, S. (1999). "Standard RGB color spaces." In "Proceedings of The Seventh Color Imaging Conference: Color Science, Systems, and Applications," pages 127–134.

Terry, M. and Mynatt, E. D. (**2002**). "Recognizing creative needs in user interface design." In "C&C '02: Proceedings of the 4th conference on Creativity & Cognition," pages 38–44. ACM Press, New York, NY, USA.

Tornquist, J. (**1999**). *Colore e Luce: Teoria & practica (in Italian)*. Istituto del Colore.

Tractinsky, N., Katz, A. S., and Ikar, D. (**2000**). "What is beautiful is usable." *Interacting with Computers*, 13(2): 127–145. URL http://dx.doi.org/10.1016/S0953-5438(00)00031-X. Accessed May 20, 2010.

Travis, D. (**1991**). *Effective Color Displays: Theory and Practice*. Academic Press (University Press, Cambridge).

Tufte, E. R. (**1990**). *Envisioning Information*. Graphics Press.

Tufte, E. R. (**1997**). *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press.

Tufte, E. R. (**2001**). *The Visual Display of Quantitative Information*. Graphics Press, 2nd edition.

Valdez, P. and Mehrabian, A. (**1995**). "Effects of color on emotions." *Journal of Experimental Psychology*, 123: 394–409.

van Geel, F. (**2006**). "Online shopping decision making: The influence of background colors and screen positioning." In "Online Proceedings of Twente Student Conference on IT - 4th Referaat," URL http://referaat.cs.utwente.nl/documents/2006_04_C-Intelligent_Interaction/2006_04_C_Geel,F.V.H.van-Online_shopping_decision_making_The_influence_of_background_colors_and_screen_positioning.pdf. Accessed May 20, 2010.

Varley, H., editor (**1980**). *Colour*. Mitchell Beazley Arts House.

Vièno t, F. and Brettel, H. (**1999**). "Digital video colourmaps for checking the legibility of displays by dichromats." *Color Research & Application*, 24(4): 243–252. URL http://dx.doi.org/10.1002/(SICI)1520-6378(199908)24:4<243::AID-COL5>3.0.CO;2-3. Accessed May 21, 2010.

von Goethe, J. (**1810**). *Farbenlehre*. Weimar.

von Goethe, J. W. (**2006**). *Theory of Colours*. Dover Publications. Reprint of 1840 Translation with notes by Charles Lock Eastlake. Original work published in 1810.

von Goethe, J. W. and Matthaei, R. (**1971**). *Goethe's color theory*. Studio Vista. Arranged and edited by Rupprecht Matthaei. English edition translated and edited by Herb Aach.

W3C (**2007**). "Understanding WCAG 2.0: a guide to understanding and implementing Web Content Accessibility Guidelines 2.0." URL http://www.w3.org/TR/2007/WD-UNDERSTANDING-WCAG20-20070517/Overview.html. Accessed May 20, 2010.

Wakita, K. and Shimamura, K. (**2005**). "SmartColor: disambiguation framework for the colorblind." In "Assets '05: Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility," pages 158–165. ACM Press, New York, NY, USA. URL http://doi.acm.org/10.1145/1090785.1090815. Accessed May 20, 2010.

Walch, M. and Hope, A. (**1995**). *Living Colors: the definitive guide to color palettes through the ages*. Chronicle Books.

Walraven, J. and Alferdinck, J. W. (**1997**). "Color Displays for the Color blind." In "Proceedings of the 5th Color Imaging Conference," pages 17–22. Society of Imaging Science & Technology.

Wang, A.-H. and Chen, M.-T. (**2000**). "Effects of polarity and luminance contrast on visual performance and VDT display quality." *International Journal of Industrial Ergonomics*, 25(4): 415–421. URL `http://dx.doi.org/10.1016/S0169-8141(99)00040-2`. Accessed May 20, 2010.

Wang, Y. D. and Emurian, H. H. (**2005**). "An overview of online trust: Concepts, elements, and implications." *Computers in Human Behavior*, 21: 105–125.

Wesner, M. F. and Shevell, S. K. (**1992**). "Color-Perception within a Chromatic Context - Changes in Red Green Equilibria Caused by Noncontiguous Light." *Vision Research*, 32(9): 1623–1634.

Westland, S., Laycock, K., Cheung, V., Henry, P., and Mahyar, F. (**2007**). "Colour Harmony." *Colour: Design & Creativity*, 1(1): 1–15. URL `http://www.colour-journal.org/2007/1/1/07101article.pdf`. Accessed May 20, 2010.

Whelan, B. M. (**1997**). *Color Harmony 2 with The Palette Picker: A Guide to Creative Color Combinations*. Rockport Publishers, book and cdrom edition.

Wills, P. (**1997**). *Working with Colour: a beginner's guide*. Hodder & Stoughton Educational.

Wolpert, D. H. and Macready, W. G. (**1997**). "No Free Lunch Theorems for Optimization." *IEEE Transactions on Evolutionary Computation*, 1(1): 67–82.

Woolman, M. (**2002**). *Digital Information Graphics*. Thames and Hudson.

Wright, A. (**1995**). *The Beginner's Guide to Colour Psychology*. Kyle Cathie.

Wright, P., Mosser-Wooley, D., and Wooley, B. (**2000**). "Techniques and Tools for Using Color in Computer Interface Design." URL `http://www.acm.org/crossroads/xrds3-3/color.html`. Accessed May 20, 2010.

Wright, W. D. (**1929**). "A re-determination of the trichromatic coefficients of the spectral colours." *Transactions of the Optical Society*, 30(4): 141–164. URL `http://dx.doi.org/10.1088/1475-4878/30/4/301`. Accessed May 20, 2010.

Wu, D.-L. and Chuang, M.-C. (**2000**). "Designing Color Schemes to Express the Expected Images of Personal Homepages." In "Bulletin of the 5th Asian Design Conference (CD ROM), International Symposium on Design Science, October 11-13, Seoul National University, Korea," .

Wyszecki, G. and Stiles, W. S. (**1982**). *Color science: Concepts and methods, quantitative data and formulae*. John Wiley & Sons, 2nd edition.

Xing, J. (**2006**). "Color and Visual Factors in ATC Displays." Technical Report DOT/FAA/AM-06/15, Federal Aviation Addministration - Office of Aerospace Medical Institute, Civil Aerospace Medical Institute Federal Aviation Administration Oklahoma City, OK 73125. URL `http://www.faa.gov/library/reports/medical/oamtechreports/2000s/media/200615.pdf`. Accessed May 20, 2010.

Yamakazi, H. and Kondo, K. (**1999**). "A Method of Changing a Color scheme with Kansei scales." *Journal for Geometry and Graphics*, 3(1): 77–84.

Zollinger, H. (**1988**). "Categorical color perception: Influence of cultural factors on the differentiation of primary and derived basic color terms in color naming by Japanese children." *Vision Research*, 28(12): 1379–1382.

Zuffi, S., Beretta, G., and Brambilla, C. (**2006**). "A Color Selection Tool for the Readability of Textual Information on Web pages." In "Proceedings of IS&T/SPIE Internet Imaging VII," SPIE.

Zuffi, S., Scala, P., Brambilla, C., and Beretta, G. (**2007**). "Web-based versus controlled environment psychophysics experiments." In L. C. Cui and Y. Miyake, editors, "19th Annual Symposium Electronic Images Science and Technology," volume 6494. SPIE. URL `http://dx.doi.org/10.1117/12.703926`. Accessed May 20, 2010.