Computational Production of Colour Harmony. Part 2: Experimental Evaluation of a Tool for GUI Colour Scheme Creation

Giovanni Moretti, Stephen Marsland, Paul Lyons*

School of Engineering and Advanced Technology, Massey University, Ag/Hort Building, Private Bag 11-222, Palmerston North, New Zealand

Received 10 October 2010; revised 25 July 2011; accepted 24 August 2011

Abstract: Although webpage and computer interface designers generally have little experience at generating effective colour schemes, colour selection appears only rarely in user interface design literature. This article describes the experimental evaluation of an algorithmic technique that applies colour harmony rules to the selection of colour schemes for computer interfaces and web pages. The technique uses a genetic algorithm to evolve colour schemes; the evolutionary path is determined by a quantitative colour harmony evaluation function.

Our technique first creates abstract colour schemes by applying those rules to specific features of the interface or web page; the user then holistically modifies the scheme's overall colour cast, overall saturation, and light–dark distribution, producing colourings that are both harmonious and usable. We demonstrate experimentally that the software is relatively simple to use and produces colourings that are well-received by humans.

In an earlier article, the criteria for a colour harmony tool for computer interfaces and websites were described and used in the design of the Colour Harmoniser, our software implementation of a system that is based on classical rules of colour harmony, adapted and extended to suit graphical user interfaces.

In this article, we describe two sets of experiments that have demonstrated the usability and effectiveness of the Colour Harmoniser tool, compared with standard methods of colour selection. These experiments suggest that the tool functions somewhat more effectively than we originally anticipated, producing colour schemes that were rated more highly on several quality scales than those produced by random choice, by humans who self-classify as nonartists, and by humans who self-classify as artists.

Col Res Appl, 00, 000–000, 2011; Published online in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/col.20735

Key words: colour harmony; Colour Harmoniser; colour scheme evaluation; automatic colour scheme generation

INTRODUCTION

In the preceding article,[1] we have presented a set of criteria that a colour harmonization tool should fulfill, and described the implementation of a prototype for such a tool. The implementation established that it is in fact feasible to construct a practical tool that conforms to those criteria, the prototype that we call the Colour Harmoniser.

In this article, we present an experimental validation of the Colour Harmoniser that evaluates the usability and effectiveness of the tool. There is considerable scope for improving the usability of colour selection tools: existing tools do not provide a mechanism for developers to colour an interface or a webpage in a holistic, direct-manipulation, real-time manner.[2] Instead, they mostly provide indirect, modal dialogs, and require the user to develop a colour scheme one colour at a time, leading to a protracted design process. By contrast, the Colour Harmoniser allows the developer of an application or a website to manipulate a small number of simple controls to change all the colours in an interface simultaneously, without disrupting its colour harmony, the visibility of any of the controls, or the readability of its text.

Two questions arise. First, does this new simple and holistic approach actually make the user's task—design of an interface colour scheme—any easier? Second, how does the quality of the colour schemes created using this new method compare with the quality of those produced by humans?

In the first article, we presented a number of colour schemes that had been produced with the help of the

*Correspondence to: Paul Lyons (e-mail: p.lyons@massey.ac.nz).

Colour Harmoniser. Those schemes are functional, in that the information content is clear (at 100% magnification) and the colour combinations conform to the classical rules of colour harmony. However, they are not works of art. How do the colour schemes produced by the Colour Harmoniser rate against those produced by other means, particularly conventional colour pickers? As the only algorithmic method we have of evaluating schemes is the Colour Harmoniser itself, which would, of course, rate its own output highly, we asked human subjects to evaluate colour schemes produced using different methods.

We describe two experiments in this article, the first to measure usability of the holistic method of colour selection (and incidentally, to produce much of the data for the second experiment) and the second to measure the effectiveness of the Colour Harmoniser colour schemes. The first one was a within-subjects design that asked participants to create two colour schemes, one using a conventional method and one using the Colour Harmoniser to select and adjust a colour scheme. Each participant then completed quality and usability questionnaires. A total of 73 subjects, 39 females and 34 males, were involved in this experiment.

The second experiment was a between-groups design and had 127 participants, with an approximately equal split between the genders and a wide variety of ages from 5 to 65+. Subjects were shown colour schemes created in five different ways (although the participants were not aware of this) and they were asked to score the scheme on four criteria using Likert scales.

As the principal aim of the Colour Harmoniser is to colour user interfaces and web pages, it was felt that an idealized or abstract image would be too unrealistic to act as a useful test environment, so a realistic, but bland, interface was used. This webpage has seventeen colourable elements: page text, page body, header text, header background, footer text, footer background, navigation panel background, and text and background for five buttons. For such an interface, it is sensible to compose four single-colour groups comprising all five button backgrounds, the button text, the header and footer backgrounds, and the header and footer text. This reduces the number of colours required to seven, and the size of the search space. Even so, the number of different colourings (assuming 24-bit colour) is 2^{24^7}, which is approximately equal to 4×10^{50}. The number of usable colour schemes with visibly different colours is much smaller than this, but nevertheless, a user presented with an RGB colour selector has an enormous number of colour combinations to choose from, and as will be shown, good colour schemes are rare.

The use of a single interface removes the type of the interface as a confounding factor, and for a given number of samples, enables more robust statistical analyses. The Colour Harmoniser has been used to create colour schemes for other types of interface (presentation slide mock-ups, and application software interfaces), with results that look broadly similar to those obtained for the webpage. We therefore feel that it is likely that the results obtained from the experiments on the webpage will be able to be generalized to the other styles of interface.

For both experimental trials, the subjects we chose were members of the general public: unpaid volunteers found in various locations. Most were passers-by and visitors to a museum and art gallery, and a design school foyer; some were found in other environments, such as University computing labs, a Linux user-group meeting, businesses and homes. Each subject provided normal demographic information (age, ethnicity, and gender), and we also asked the subjects about their colour design background, but did not screen them for Colour Vision-Deficiencies (CVD), although they were asked to state—if they knew—whether their colour vision was normal or impaired.

It could be argued that web design is an activity best left to professionals and that professional designers should have been used as the experimental subjects, particularly for the design experiment. However, many millions of amateurs create a variety of web pages that are based on templates but can be personalized in different ways—blogs, and pages on media such as Twitter and Skype and MySpace. This group has a much greater need for assistance with colour choice than professionals but in general, the free software to which they have access provides poor tools for making colour choices—often nothing more than a small selection of ready-made palettes or themes. Our system is directed at such users, and we therefore felt that it was appropriate to choose subjects from the general public for both the design and evaluation experiments, rather than professional designers.

This same argument explains why we did not exclude CVD participants, as they are users of both websites and software to create them. An important factor in the usability of interfaces is text legibility, and the Colour Harmoniser ensures legibility by emphasizing lightness difference rather than hue difference between text and its background. In fact, there were very few CVD participants, and no CVD effect in the Harmoniser-based trials. Colour schemes created by participants with known CVD were omitted from the trials in which human-created schemes and Colour Harmoniser-created schemes were compared. An additional reason not to test for CVD, but simply to ask participants if they knew of any impairment was to ensure the participants completed the experiment, which was already reasonably complex and time intensive, consisting of the five tasks detailed in the next section, each task having each with multiple screens.

The environments where the experiments took place were uncontrolled, and typical of those used by home users and nonprofessional 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. The type of lighting was noted in each case.

The images used in the experiments were presented on modern LCD panels adjusted to a luminance of 120 cd/m^2 on machines running Windows XP™, with the system colour profile set to an ICC sRGB colour profile created by a GretagMacbeth EyeOne-Pro™ spectrophotometer. As noted by Laugwitz,[3] colour harmony experiments under more controlled experimental conditions could be expected to yield similar results.

USABILITY OF THE COLOUR HARMONISER

This experiment had two aims: to compare the Colour Harmoniser's usability as a colour-selection tool (as evaluated by the participants) with that of conventional "one-colour-at-a-time" methods, and to compare the times required to create colour schemes using the two methods. A side-benefit was that the colour schemes produced during the course of this usability experiment were available to be used in the subsequent effectiveness experiment.

Subjects created two colour schemes for the webpage, one using the Colour Harmoniser and the other using a conventional method. The order of the two tasks was varied to remove learning effects. The trial was computerized and involved five stages:

1. an introduction that explained the purpose and structure of the experiment, and the identity and organizational affiliation of the experimenter;
2. a demographic survey, in which the subjects provided their gender, age, artistic self-perception, and colour training. Each question had a popup hint expanding on the question or giving a rationale for its inclusion, and all questions were optional;
3. the creation of two colour schemes, one by the conventional method (described below), and one using a Colour Harmoniser-created scheme;
4. a post-creation questionnaire after each scheme creation task, with questions related to the usability of the tool and the quality of the resulting colour scheme;
5. a concluding survey in which the participants compared the usability of the two tools, noted their satisfaction with each of the schemes and made a direct comparison between the quality of their two resultant schemes.

No time limit was placed on the tasks, but times were recorded. No information about the identity of the participants was recorded. The experiments were approved by the Massey University Human Ethics Committee protocol, reference #04-184.

Although, in stage 5, we collected data about the participants' satisfaction with the two colour schemes, we have omitted it from the results presented here, as we were concerned that it could be contaminated by a sense of ownership of the scheme that they had produced from scratch (the one in which they chose all the colours individually). The effectiveness experiment presented below gives more robust results for the evaluation of colour scheme quality, as participants in that experiment did not know how schemes were created, and were not personally involved in their creation.

The software created for the experiment was designed to allow the user to focus on the colouring task, without being overwhelmed by irrelevant detail such as selection methods and navigation. Simultaneous contrast effects were minimized by presenting the test interface on a mid-gray background. Custom user interfaces were created for each of the colour scheme creation tasks. Both are described below, with screen-shots of the stages during each colouring. In practice, the participants found both colouring interfaces easy to work with and required little in the way of additional explanation. The survey questions were also designed with simplicity in mind, and additional detail in the form of pop-up hints was provided.

To avoid the use of the loaded terms "Colour Harmoniser" and "Manual Method," the two approaches were referred to as "All Colours Together" and "One Colour at a Time," respectively, the former a reference to the Harmoniser's holistic method of colour scheme adjustment, the latter to the more traditional method of creating a colour scheme by selecting colours individually for each interface object.

Experimental Procedure (All Colours Together Scheme Creation)

The interface for the Colour Harmoniser has two screens, a "home" screen and the holistic colouring adjustment interface, which is invoked when required. The home screen displays a set of potential "starter" schemes created by the Colour Harmoniser prototype, in thumbnail form.

Ten "starter" schemes were presented to each user. These were created by the Colour Harmoniser with the seven colour groupings described earlier, and the following were constrained to be mutually distinguishable (i.e., to use visibly different colours): the body background colour; the header/footer background group; the button background group; and the left hand navigation bar background.

For the creation of the colourings, the wireframe was titled at 45° to the horizontal, but hue rotation was chosen randomly. Eighty schemes were generated, 20 each of monochromatic, analogous, complementary (10 including colours on the black-white axis, 10 not), and split-complementary (10 including colours on the black-white axis, 10 not). Each user saw 10 starter schemes randomly selected from this set of 80.

Figure 1 shows an example of the screen on which thumbnails of these starter schemes were presented. In addition to the thumbnail region, the screen incorporates two larger areas onto which thumbnails can be dragged. There they are displayed at a larger magnification, to make it easier for the user to make detailed comparisons of subtle colour differences between the images.

Users could choose to modify either of those two schemes, which would cause the colour adjustment screen shown in Fig. 2 to be displayed. The chosen scheme is shown full-size, at the top of the window. Users could

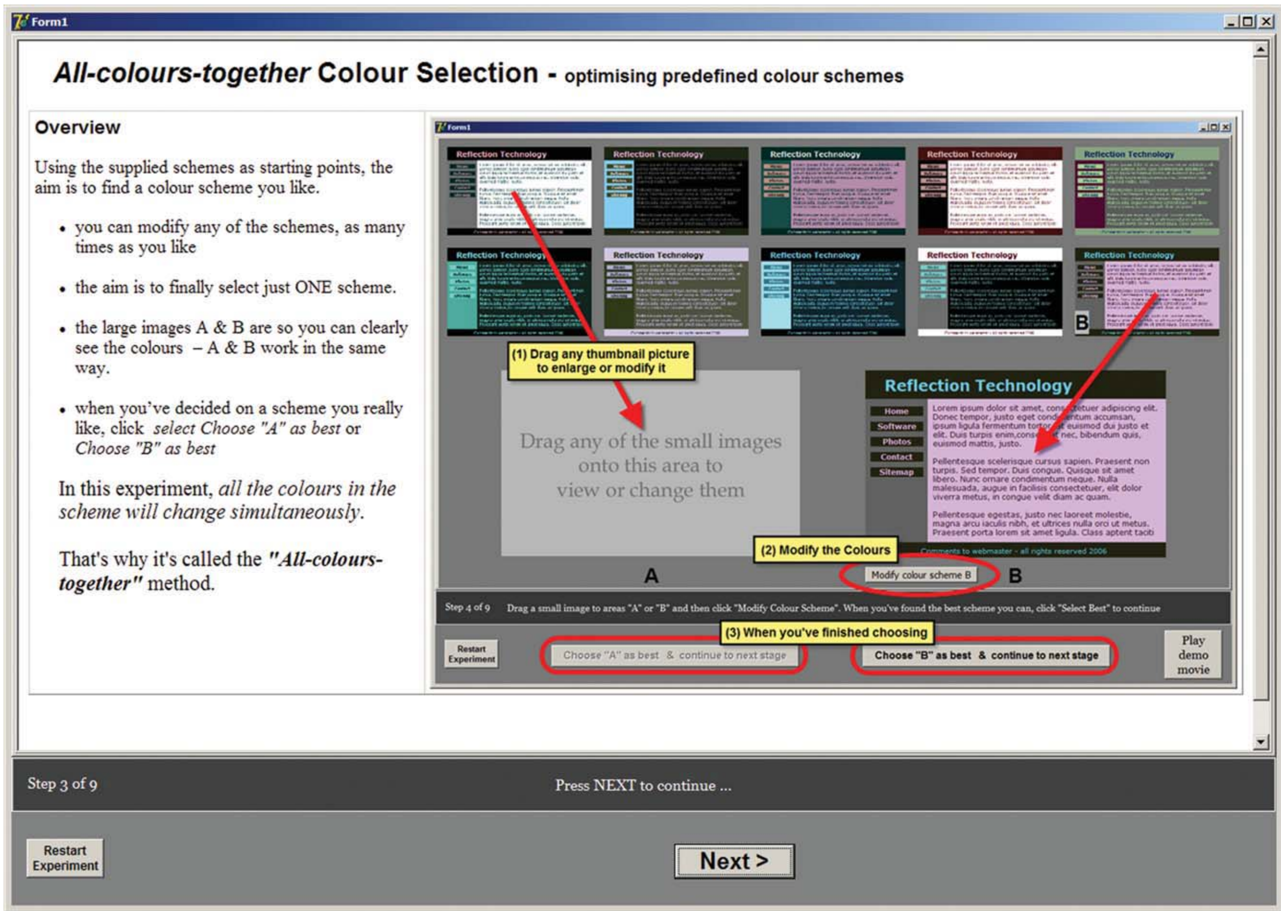

FIG. 1. The introductory screen of the primary interface for the all-colours-together (Colour Harmoniser) task informs the participant that any 2 of the 10 screens in the top of the window can be dragged to the fields at the bottom for comparative purposes, and that there is a button to click to modify the colours.

holistically modify the scheme's hue rotation, saturation, and light–dark distribution using the controls at the bottom of the screen. Once they were satisfied with their changes, they could return to the first screen, preserving the changes in both the enlarged scheme on the home page and its related thumbnail. Returning to the home page without saving changes is also possible. The user could modify as many or as few schemes as often as they liked, compare them in pairs before choosing one as their final preferred colouring.

Figure 3 shows two colour schemes that have resulted from user modifications of the scheme shown in Fig. 2. In the left scheme, the user has used the flip light–dark control to make light colours dark, and dark colours light. In the scheme on the right, further modifications have occurred; a hue rotation has affected all the colours in the interface and their saturations have been increased.

Experimental Procedure (One Colour at a Time Scheme Creation)

As a control against which to compare the Colour Harmoniser method of colouring we also presented participants with a custom interface intended to typify common colour selection interfaces (Fig. 4), without requiring the subjects to have prior familiarity with any existing application.

This interface uses a conventional colour picker for colouring the same interface as was used before. When the user clicks on any item in the interface, the colour picker

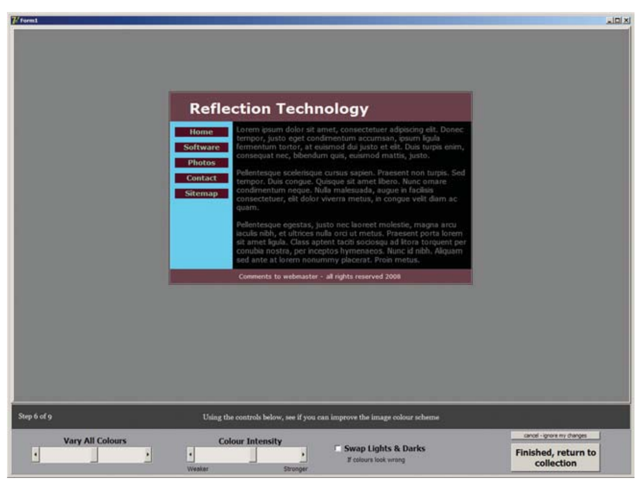

FIG. 2. The secondary interface for the Colour Harmoniser task provides simple slider controls for holistic hue rotation, saturation alteration and lightness inversion.

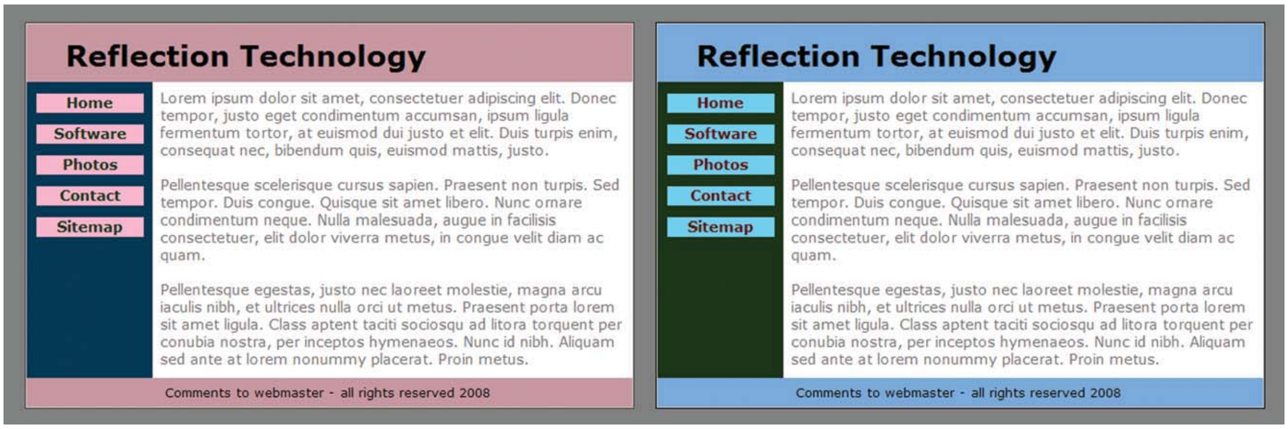

FIG. 3. Two possible outcomes from the participant's tweaking of the raw colour scheme shown in Fig. 2. In the first, the only change is that the lightness has been inverted. In the second, the participant has also altered the hue rotation and increased the saturation. These images are shown side-by-side here to economise on space: in the experimental setup, they would be displayed sequentially in the same window.

is displayed and the click-target is selected, along with the other members of any group of identically coloured items to which it belongs. Then, when the user has selected a colour with the colour picker, all of those items are coloured. Textual areas are treated slightly differently, as clicks on text often miss their target and hit the background, making text selection difficult. Therefore, when the user clicks anywhere on a button or the body text, a pop-up menu appears with the options *Change colour of words* and *Change colour of background* (or buttons, as appropriate), so that the user can choose the interface component to be coloured. A single demonstration was sufficient to teach the participants how to use this interface.

Figure 5 shows the standard Windows™-style colour picker that appears once a participant has clicked on the interface item to be coloured. At this point, the familiar iterative colour-selection process occurs; the user calls up the colour picker, selects the colour for the selected item (or group of items), and closes the colour picker. This process is repeated for each item in the interface until the overall colour scheme was found to be acceptable, possibly with repetitions when a colour is found to be incompatible with other colours.

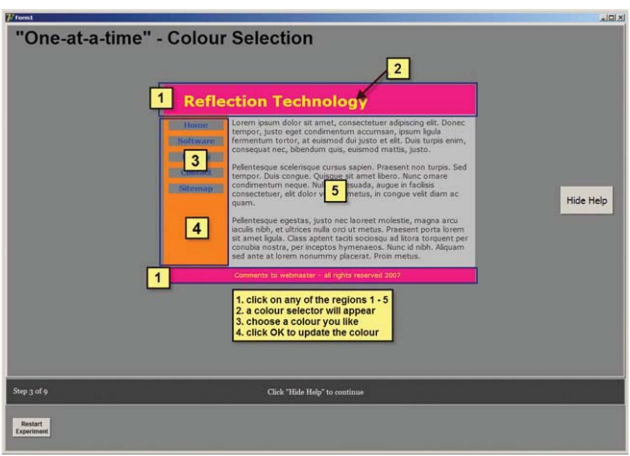

FIG. 4. The interface for choosing the items to be coloured (numbered 1–5) using the manual colouring approach allows a user to alter the colours of identically coloured items (such as the header and footer backgrounds) at once. For the buttons (3) and the body (5), both the text and the background colour may be altered.

Post-Colouring Surveys

The order of the two colouring tasks was chosen at random for each participant, and participants were asked to complete a questionnaire after each task to discover how usable they found the approach. Participants could decline to answer any question.

For the holistic approach, participants were asked the questions in Table I.

For the manual (one colour at a time) approach the participants were asked the questions in Table II. Note that the last question is clearly not one that we expected members of the general public without colour design experience to be able to answer. It is included so that we could gauge how those that do consider themselves to be colour experts actually choose colour schemes in practice. The "do not know" and "skip this question" options were available for anybody who did not understand what was being asked.

After participants had completed both colouring tasks, they were shown both schemes side-by-side (in randomized order) and asked to *Click on the image with the more professional colour scheme*. They were then asked to fill out a post-task survey that included the questions in Table III.

Comparing Methods of Scheme Creation: Results

We present some selected highlights of the experimental results in this article. A more extensive analysis and further discussion are presented elsewhere.[2]

The results were analyzed using the binomial test and Student *t*-test. The number of participants who answered each question is recorded in the degrees of freedom (*df*) report. The judgements appear to be unaffected by the

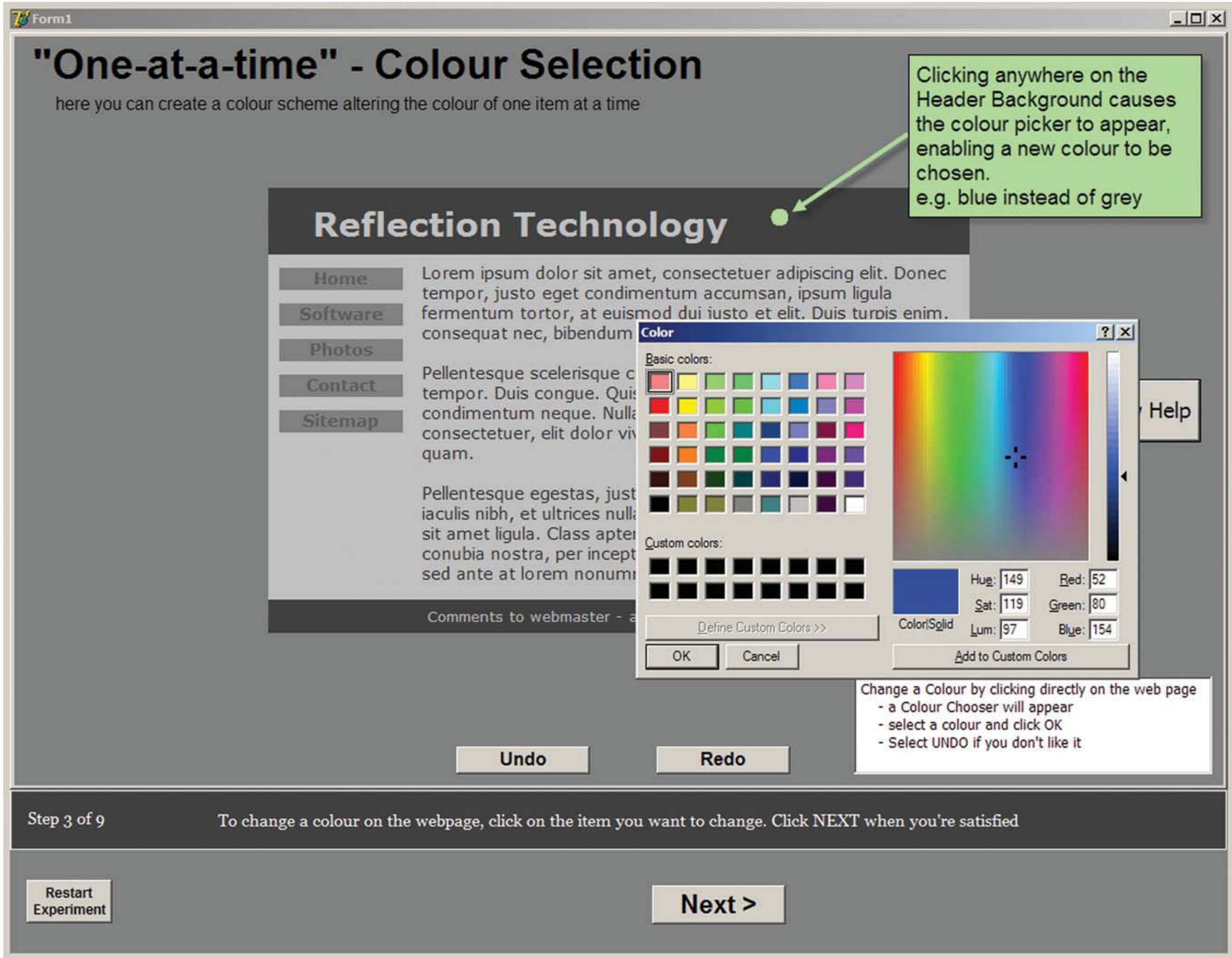

FIG. 5. A standard Windows™-style one-colour-at-a-time colour selection dialog is used in the manual colouring task.

gender or colour sense of the evaluator, or by previous colour training.

Interestingly, while there was no significant difference between the times to complete the task using the two interfaces, participants believed that it would be quicker to find a professional colour scheme using the Colour Harmoniser (binomial test 30/40, $p = 0.0022$). Most participants took less than 5 minutes to colour the interface using either scheme. Nor was there any difference in users' opinions of which system helped to produce more professional schemes.

When asked about the Colour Harmoniser, participants indicated that:

- it was reasonably easy to find a scheme they liked (one-sided t-test: $t = 1.65$, $df = 70$, $p = 0.052$)
- it was easy to understand the effect of the vary all colours and colour intensity controls (one-sided t-test: $t = 13.29$, $df = 71$, $p = 0.001$)
- it was possible to improve the schemes using the holistic controls
- it was beneficial to use the *flip-light-dark* control
- finding a good colour scheme would be easier using the Harmoniser. The difference between the mean value (3.7) and the null hypothesis value 3 ("about the same") is statistically significant (t-test: $t = 5.03$, $df = 71$, $p = 0.001$).

TABLE I. Survey questions for qualitative evaluation of the holistic (all colours together) approach.

How easy was it easy to find a colour scheme you liked?	Very easy|easy|not sure|difficult|very difficult
How much could you improve the supplied colour schemes?	Significantly|a moderate amount|a little|not at all
Did the 'Flip Light/Dark' option improve colour schemes?	Significantly|a moderate amount|a little|not at all
Ten colour schemes were provided. Would you have preferred:	Far fewer|fewer|ten was about right|a few more|many more
Was the effect of 'Vary All Colours' and 'Colour Intensity' controls easy to understand?	Very easy|easy|not sure|difficult|very difficult
How would you rate the colour scheme after your adjustments?	Professional|better than average|average|not very good|terrible

TABLE II. Survey questions for qualitative evaluation of manual (one colour at a time) approach.

Was choosing a colouring scheme?	Much easier than I expected|easier than I expected|about what I expected|more difficult than I expected|much more difficult than I expected
Was it easy to find a particular colour?	Very easy|easy|not sure|difficult|very difficult
Was it easy to keep the text readable?	Yes|no
How satisfied are you with this scheme?	Very satisfied|satisfied|neither satisfied nor dissatisfied|dissatisfied|very dissatisfied
Is your colour scheme appearance?	Professional|better than average|average|not very good|terrible
Is your colour scheme based on...?	Monochromatic colours|analogous colours|complementary colours|split complementary colours|colour triads or tetrads|an elliptical group of colours|a scheme not listed above|do not know

From prior experience, it was expected that when picking colours individually using the manual method, two aspects of the task would cause difficulty: choosing appropriate colours while keeping the text readable, and finding particular colours. The first expectation was confirmed; 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 manually would be difficult, was not supported. The mean score in answer to the question "Was it easy to find a particular colour?" was 3.4. This indicates a small but statistically significant (*t*-test: $t = 2.74$, $df = 71$, $p = 0.0078$) bias toward the "easy" end of the scale (where *very difficult* scored 0, *not sure* scored 3, and *very easy* scored 5).

After using both approaches and being asked to compare the two, participants stated that they thought the Colour Harmoniser

- made it easier to find good looking schemes,
- was more likely to yield professional looking colour schemes,
- had colouring controls that were easy to understand, and
- produced results that were comparable in quality to colour schemes produced manually.

These results are very positive. Although the time taken to create the schemes with the Colour Harmoniser was no less, the users expected the results to be more professional. The next, larger scale, experiment, in which participants scored colour schemes without being aware of their origins, enabled us to test these expectations free from the possibility of owner-bias.

Effectiveness of the Colour Harmoniser

The previous experiment compared the ease of producing colour schemes using the Colour Harmoniser with a more conventional method of colour scheme creation. Our second experiment measured the relative quality of colour schemes produced by the Colour Harmoniser and several other methods.

This experiment was a between-groups design with a dataset comprised of a number of differently coloured images of the webpage shown earlier. These had been created in the earlier usability experiment: by subjects who self-classified as artists (48 schemes); by subjects who self-classified as nonartists (93 schemes); by the Colour Harmoniser without tweaking (the 80 schemes that were created for the usability experiment); by the Colour Harmoniser, with adjustment (the 62 tweaked schemes produced during the usability experiment); and by randomly colouring the elements, with the element groupings specified in the previous experiment (100 schemes). Schemes from known CVD subjects were excluded from the artist/nonartist pool.

Each subject (subjects who had participated in the first experiment were excluded from the second experiment to avoid "ownership" of a colour scheme as a confounding factor) evaluated fifteen images, three from each of the five classes chosen randomly, and presented in a random order. The images were selected afresh for each subject, so all participants saw different image sets. Participants were asked to rate each of the schemes they saw for visual appeal, professionalism, artisticness, and suitability for a website, each on a five-point Likert scale.

As before, this experiment was conducted on a computer, with a custom-designed survey interface. The survey first asked for demographic details such as age, gender, and self-assessed artistic talent. Then the user was shown each of the fifteen images individually. Figure 6 shows the appearance of this screen, with the webpage image taking up the top two-thirds of the screen, and the four survey questions and their Likert options presented below.

Assessing Colour Harmoniser Scheme Quality: Results

There were 127 participants in total, from various backgrounds: members of the public, design students, engi-

TABLE III. Post-task survey questions following both colouring tasks.

Compared with 'One-at-a-time' finding a good scheme using 'All-colours-together' is:	Much easier|easier|about the same|harder|much harder
Which selector was less frustrating to use?	The 'All-colours-together' colour selector|not sure|the 'One-at-a-time' colour selector
It would be quicker to find a professional looking colour scheme using:	The 'All-colours-together' colour selector|not sure|the 'One-at-a-time' colour selector

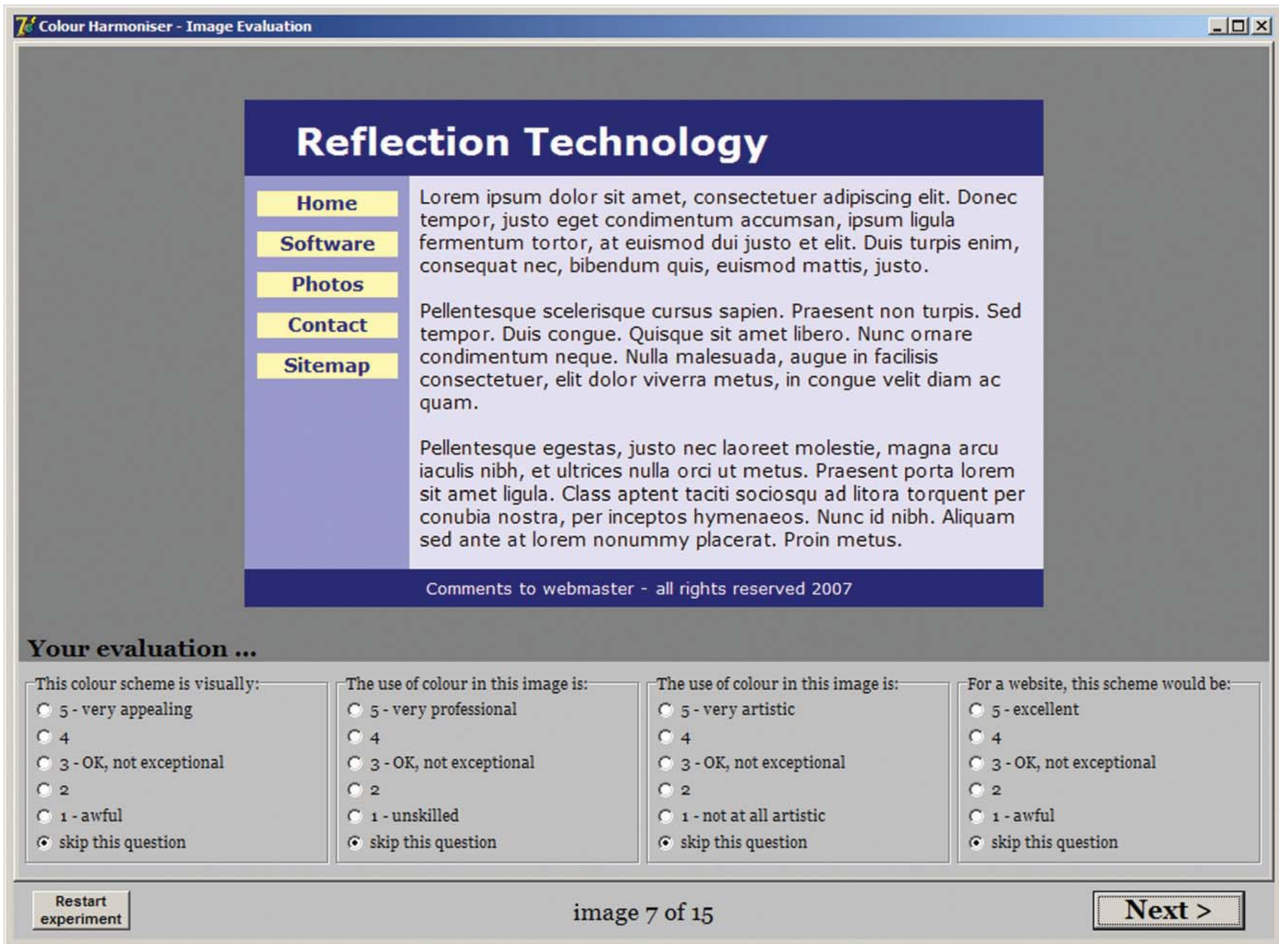

FIG. 6. The colour scheme evaluation page used in the 'Compare the Results' experiment.

neering and other nonartistic students, in a variety of locations, as described earlier.

The final data set had 1819 discrete evaluations (not all subjects completed all the evaluations, giving around 360 for each creation method). The demographics are shown in Table IV. Because the participants were free to skip questions, the totals do not always sum to 1819.

Table V shows the ratings that the participants gave the schemes on 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.

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 *artistic evaluation*.

Table V shows that there is a clear ordering to the mean scores, and the Colour Harmoniser schemes compare well against those produced by other methods, although not all of the differences are statistically significant (details in Table VI). 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 nonartists. From the mark-

TABLE IV. Demographics of survey participants.

Gender		Female 904 (50%)			Male 900 (50%)		
Artist talent (self-assessed)		Artists 603 (35%)			Nonartists 1127 (65%)		
Age	5–14	15–24	25–34	35–44	45–54	55–64	65+
	2%	28%	18%	14%	19%	13%	6%
Colour training		Yes 39%			No 61%		
Colour sense (self-assessed)	Terrible	Not very good	Average	Better than average		Excellent	
	0, 0%	133, 7%	876, 49%	660, 37%		105, 6%	
Colour vision		Normal colour vision (1744, 97%)			Impaired vision (60, 3%)		

TABLE V. Mean evaluation scores for colour schemes produced by different methods (1 is worst, 5 is best).

Source of scheme	Evaluation scores							
	Visual appeal		Professionalism		How artistic		Web suitability	
Harmoniser adjusted	$\mu = 3.15$	$\sigma = 1.01$	$\mu = 3.20$	$\sigma = 1.07$	$\mu = 2.90$	$\sigma = 0.97$	$\mu = 3.12$	$\sigma = 1.10$
harmoniser raw	$\mu = 2.91$	$\sigma = 1.10$	$\mu = 2.89$	$\sigma = 1.12$	$\mu = 2.78$	$\sigma = 1.08$	$\mu = 2.80$	$\sigma = 1.18$
Human artists	$\mu = 2.83$	$\sigma = 1.24$	$\mu = 2.77$	$\sigma = 1.19$	$\mu = 2.78$	$\sigma = 1.17$	$\mu = 2.75$	$\sigma = 1.24$
Human nonartists	$\mu = 2.80$	$\sigma = 1.23$	$\mu = 2.72$	$\sigma = 1.18$	$\mu = 2.67$	$\sigma = 1.16$	$\mu = 2.64$	$\sigma = 1.25$
Random colouring	$\mu = 1.74$	$\sigma = 0.95$	$\mu = 1.73$	$\sigma = 0.91$	$\mu = 1.83$	$\sigma = 1.00$	$\mu = 1.56$	$\sigma = 0.84$

edly low scores of the randomly coloured schemes, it is clear that schemes created using this method are less acceptable (1 corresponds to "the *awful*" or "*without merit*") end of the scale.

It is easy to create colour schemes that look really bad, but much harder for them to look extremely good, and this is reflected in the mean scores shown in Table V. The differences between the mean 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 nonartistic), 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 (Fig. 7). A larger version of these plots is available online.[2]

To enable an evaluation of the merit of the adjusted Colour Harmoniser-derived schemes, Table VI shows the significance of the differences between the means of the adjusted Harmoniser-based colour schemes and those created using other methods. 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.

The overall impression to be gained from these results is that the Colour Harmoniser has significantly exceeded our original expectations. It produces results that are visually appealing, appear professional, and would be suitable for use online.

CONCLUSIONS

The impetus for this research was the observation that designing good user interface colour schemes is difficult and the colour selectors in most applications do little to ameliorate this.[4] Although most widely used applications allow changes to the colour of individual items such as text, shapes and backgrounds, they provide no support for ensuring that the overall effect is harmonious. This is a little surprising, as heuristics for the creation of pleasing colour schemes are well-known in the artistic community.[5–7] 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[8] but as these guidelines and models are intended to aid artists and graphic designers, factors pertinent to interfaces and usability are not included.

In these articles, we have investigated 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. Our model considers an interface colour scheme as a whole, taking account of the relationships between its components during its creation and modification, by using the idea of a rotatable wireframe, a concept that underlies many of the simpler models of colour harmony.

An extended model of colour harmony based on these ideas was developed and embedded in a piece of software called the Colour Harmoniser, which can analyze the structure of existing interfaces and create optimized abstract colour schemes. The extended 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 optimized abstract scheme.

In this article, we have presented a series of experimental trials and results that validate the extended model of

TABLE VI. Significance of the differences between the means of the Harmoniser-created colour schemes after human adjustment and those created using all other methods.

	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	$p < 0.001$***	$p < 0.001$***	$p = 0.121$	$p < 0.001$***
Nonartist	$p < 0.001$***	$p < 0.001$***	$p = 0.004$**	$p < 0.001$***
Random	$p < 0.001$***	$p < 0.001$***	$p < 0.001$***	$p < 0.001$***

Statistical significance for: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
The order of the means, from least to greatest is: random, nonartist, artist, Colour Harmoniser (raw), Colour Harmoniser (adjusted).

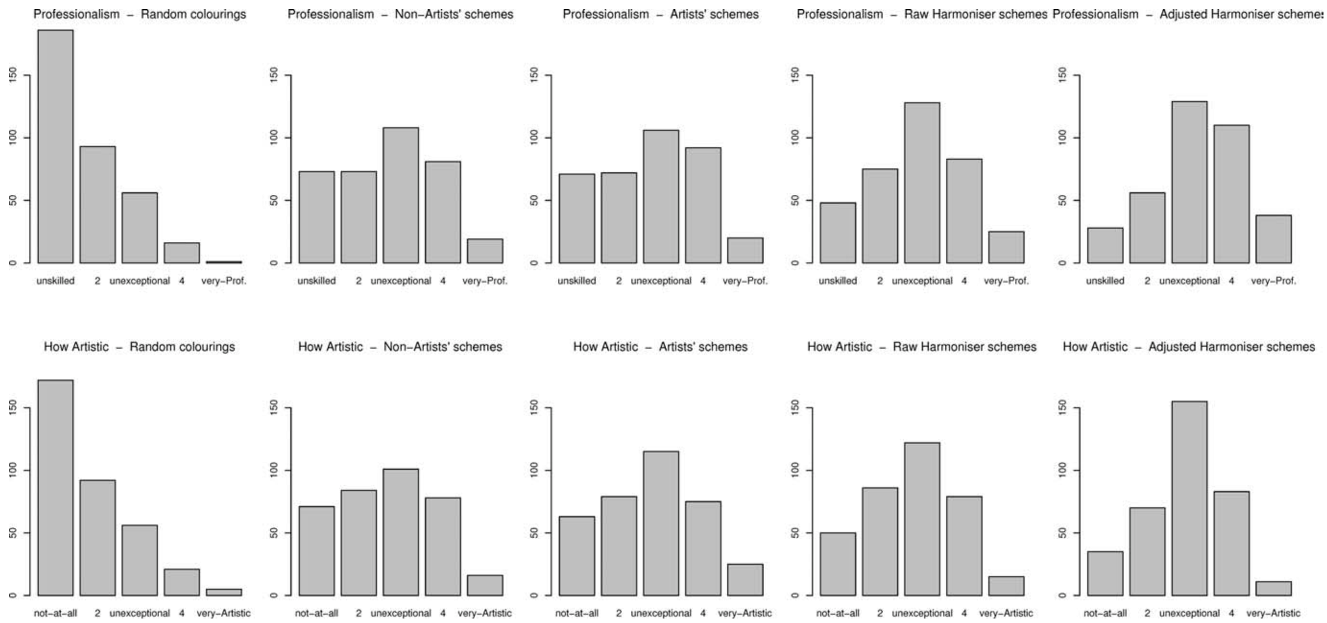

FIG. 7. A Histogram showing the participants' evaluation of how professional (upper row) and how artistic (lower row) the colour schemes were for each creation method. The categories within each plot indicate increasing quality from left-to-right: unskilled to very professional (top row), and not-at-all to very artistic (bottom row). From left to right, the columns of plots indicate schemes coloured: randomly, by a human nonartist, human artist, Harmoniser-raw, and Harmoniser-adjusted. There is a clear improvement in quality from left to right. A larger version of these plots is available elsewhere.

colour harmony, as exemplified by the Colour Harmoniser's fitness function, and compared 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.

The results demonstrate that the proposed method of creating user interface colour schemes by combining colour harmony heuristics with constraints related to usability and user interface semantics is both viable and robust. Specialist knowledge of colour harmony is not required, either to create schemes, or to personalize them once created. The colour schemes produced using the extended model of colour harmony are seen as visually appealing, professional and suitable for online use across a wide cross section of users. The experimental results broadly validate the saturation/area-balance and lightness/area-balance models of harmonious colour scheme design. The poor user evaluations of randomly coloured schemes also provide experimental evidence for the relative scarcity of good colour schemes in the colour scheme design space. 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.

Heartening though those results are, it is not claimed that those rules are optimal; they could be improved, augmented, or even replaced. For example, different sets of rules could be used to generate website colour schemes for young children, allowing for their different aesthetic preferences or, if required, to create discordant nonharmonious colour schemes to evoke feelings of unease, while in both cases ensuring the usability of the created schemes and allowing them to be personalized.

We have made several simplifying assumptions in our research to date, which has acted as a proof of concept. There are a few areas that we believe are of importance for future investigations, such as allowing for interfaces that include texture, gradation and transparency, maintaining colour harmony over multiple pages, incorporating fixed-colour content, such as photographs and figures, and limiting readability and distinguishability contrast (the current approach trades off subtlety in the colouring for guaranteed readability/distinguishability in the final colouring). Additionally, the current architecture is intended to minimize the requirement for up-front specification and to allow pleasing colour schemes to be discovered by exploration as the user holistically adjusts 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, such as allowing the nonholistic adjustment of colours elements within a scheme as an occasional exception to the colour rules, possibly with oversight by a module using the fitness function that alerts the user if the changes are severely detrimental.

ACKNOWLEDGMENTS

This work forms part of a PhD thesis in the School of Engineering and Advanced Technology at Massey University. There was substantial help in the experimental design and statistical analysis from Alasdair Noble, Jonathan Godfrey, and Geoff Jones, and their help is gratefully acknowledged.

1. Moretti G, Lyons P, Marsland S. Computational Production of Colour Harmony. Part I: A Prototype Colour Harmonization Tool. Color Res Appl 2011, DOI 10.1002/col.20736.
2. Moretti GS. A Calculation of Colours: Towards the Automatic Creation of Graphical User Interface Colour Schemes. PhD Thesis. Palmerston North: Massey University; 2010. 280 p. Available at: http://hdl.handle.net/10179/1492.
3. Laugwitz B. A web experiment on colour harmony principles applied to computer user interface design. In: Reips UD, Bosnjak M, editors. Dimensions of Internet Science. Lengerich, Germany: Pabst Science Publishers; 2001.
4. Lyons P, Moretti G. Nine tools for generating harmonious colour schemes. In: Masoodian M, Jones S, editors. Proceedings of 6th Asia Pacific Conference on Computer Human Interaction (APCHI 2004). Rotorua, New Zealand: Springer-Verlag; 2004. p 241–251.
5. Albers J. Interaction of Color: Revised and Expanded Edition. New Haven, CT: Yale University; 2006. 145 p.
6. Caivano JL. GAC Revista: Grupo Argentino del Color. 19 April, 2004. Report nr ano 7, no19. 20 p.
7. Itten J. The Art of Color. New York: Wiley; 1970.
8. Birren F. MUNSELL: A Grammar of Color. New York: Van Nostrand-Reinhold; 1969. 80 p.