NRC Publications Archive
Archives des publications du CNRC

Preferred surface luminances in offices, by evolution
Newsham, G. R.; Marchand, R. G.; Veitch, J. A.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

NRC Publications Record / Notice d'Archives des publications de CNRC:
https://nrc-publications.canada.ca/eng/view/object/?id=15967e29-a39e-49ac-9b0d-e36bb4ff8bf2
https://publications-cnrc.canada.ca/fra/voir/objet/?id=15967e29-a39e-49ac-9b0d-e36bb4ff8bf2

Questions? Contact the NRC Publications Archive team at
PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada   Conseil national de recherches Canada

Canada

National Research
Council Canada

Conseil national
de recherches Canada

NRC·CNRC

Preferred surface luminances in offices, by evolution

Newsham, G.R.; Marchand, R.G.; Veitch, J.A.

NRCC-46976

http://irc.nrc-cnrc.gc.ca/ircpubs

construction

Institute for
Research
in Construction

Institut de
recherche
en construction

IRC

Preferred Surface Luminances in Offices, by Evolution

Guy R. Newsham, Roger G. Marchand, Jennifer A. Veitch

Institute for Research in Construction

National Research Council Canada

Ottawa, ON, K1A 0R6

guy.newsham@nrc-cnrc.gc.ca

Abstract

40 participants viewed a series of greyscale images of a typical non-daylit, open-plan partitioned office, and rated them for attractiveness. The image was projected onto a screen at realistic luminances and 54% of full size. The images in the series were geometrically identical, but the luminances of important surfaces were independently manipulated. Initially, the combinations of luminances were random, but as the session continued a genetic algorithm was used to generate new images that generally retained features of prior images that were rated most highly. As a result, the images presented converged on an individual's preferred combination of luminances. The results demonstrated that this technique was effective in reaching a participant's preferred combination of luminances. There were significant differences in room appearance ratings of the most attractive image compared to other images, and the differences were in the expected direction. Factor analysis of ratings of the most attractive images revealed a factor structure with some similarity to that obtained when people rated real office spaces. Furthermore, preferred luminances were similar to those chosen by people in real settings, as was the variation in preferences between individuals. Finally, subjective ratings of brightness, uniformity and attractiveness were significantly related to luminances in the image.

Introduction

The traditional method of exploring preferred luminous conditions involves participants evaluating full-scale physical mock-ups of spaces lit in different ways. While high in external validity (the fit between the experimental condition and a real world setting), these studies are expensive, especially if one wishes to manipulate the lighting design between evaluations. This is not only a drawback for the

1

researcher: lighting manufacturers and designers who wish to present design solutions to their clients are often required to create expensive physical mock-ups.

Partly as a response to this, there has been some interest in other, cheaper, presentation methods, such as scale models, photographs, or renderings from computer simulation packages. Research in areas such as forestry, architecture and urban design[1-3] have established that images can be a reasonable surrogate for the real space, particularly on ratings related to aethestics. The limited research in lighting on this topic concurs with this, when representing the real space with photographs[4], or with highly-detailed simulations[5].

However, these studies have been limited to the evaluation of a limited set of predefined luminous environments. With this approach one can compare ratings of images to real spaces, find which of a set of images is most preferred, and look for general trends, but one cannot easily find the optimal luminous environment. Johnston[6] and Johnston & Franklin[7] described an interesting method using computer-generated images of faces to arrive at an optimally attractive face. The software initially presented a series of faces with random variations of features (e.g., hair colour, size of chin, separation of eyes). The participant rated each of these faces in terms of attractiveness. Using a genetic algorithm, the software then combined the most attractive faces to produce new combinations of faces and the rating process was repeated until a face with an optimal rating for that participant was arrived at. The genetic algorithm proved to be a very effective method of arriving at an optimally attractive face from a vast combination of possible faces. Furthermore, the features of the preferred faces correlated well with the preferred features from other human factors studies using real stimuli. In this study we applied Johnston's method to computer-generated images of lit scenes.

This is not the first study to apply genetic algorithms to find optimal solutions to lighting problems. Ashdown[8] described a process for using genetic algorithms in non-imaging optics to find optimal luminaire designs. Eklund & Embrechts[9] used genetic algorithms to optimize filter design to develop energy-efficient light sources with desired spectral output. Chutarat & Norford[10] described an inverse method utilising genetic algorithms to derive the physical parameters of a room to produce desired

daylighting performance. However, all of these studies used user-defined physical performance criteria, not a criterion based on subjective evaluation.

Other inverse methods using deterministic optimization techniques, rather than genetic algorithms, have been applied to illumination in the computer graphics domain. Kawai et al.[11] presented a system whereby a user could specify certain target luminous conditions and the optimal luminaire focussing and output would be generated. Schoeneman et al.[12] allowed the user to "paint" lighting patterns on a rendered scene and then determined the light outputs and colours from a given set of fixed luminaires that would most closely match the desired pattern. One drawback of these automatic optimization techniques is that they are based on the user's pre-existing biases towards a desired solution and on pre-programmed weightings of various performance parameters, thus reducing the exploration of novel solutions. The genetic algorithm approach does not assume that participants can describe their preferred lighting conditions in advance of the experiment, rather that when presented with images they will know their preferred conditions when they see them. This is particularly valuable when working with naïve observers.

Moeck[13] developed a software tool to directly manipulate the luminance and chromaticity of certain surfaces in a computer-generated image. These surfaces served as light sources themselves with realistic inter-reflections allowing for the exploration of luminous patterns independent of specific luminaires. The concept is very similar to the method we use in this study, except that Moeck's software tool was designed for trial-and-error exploration as a teaching tool, and not as an optimization tool for research.

To sum up, other researchers have explored using images as a surrogate for real spaces, and have explored different techniques to find optimal lighting solutions. To our knowledge, ours is the first study to apply genetic algorithms to directly manipulate the image's luminous field experienced by observers, and to use the subjective ratings of those observers as the performance criterion rather than calculated physical parameters.

In 2001 we carried out a pilot study[14] using software to present images for rating by participants on an attractiveness scale. Participants in this pilot study were 22 lighting experts. Successive images presented to a participant varied only in the luminance of certain surfaces. The software used a genetic algorithm to develop the optimally attractive combination of surface luminances for each participant. The results were very encouraging, indicating that the method was efficient in producing attractive images, and that the preferred luminances chosen were consistent with preferred luminances chosen in experiments conducted in real spaces. Following analysis and feedback from this pilot study, the software and procedure were modified slightly. The experiment described in this paper is a replication of the pilot study (with the slightly modified software and procedure) using a larger sample of naïve participants.

This experiment was designed to test the following hypotheses:

- The genetic algorithm is an efficient method to generate a highly attractive image

- Highly attractive images are rated differently than non-optimal images

- Images are perceived in the same way as real spaces (with respect to luminance patterns)

- Preferred luminances and ratios derived from images are the same as those derived for experiments in real spaces

- Subjective ratings correlate with photometric descriptors

- The genetic algorithm method is repeatable within subjects

Methods & Procedures

The Image

We chose to use an image of an open-plan, partitioned office space (a "cubicle"). We did this for a number of reasons. These spaces are the single most common work space in North America, and therefore of importance and common to the experience of many people. They offer several simple and easily isolated surfaces to be manipulated. We are also experienced in conducting lighting experiments in real spaces of this kind. We used a photograph of a cubicle in the same physical space as that studied by Veitch & Newsham[15-16] because their results could then be compared to the results of this study.

The photograph was taken from the entrance to the cubicle, so that various important surfaces (partitions, desktop, distant boundary wall, ceiling) each occupied a reasonably large area. We wanted to eliminate harsh shadows and illuminate each surface as evenly as possible in order to reduce associations with particular light sources. For the same reason, we did not want any luminaires to be visible in the photograph. By experimentation, we found the best light source to achieve this was the camera's own flash. Unfortunately, this meant that absolute luminances were not measurable. The photograph was taken with a Kodak DC260 digital camera, in high-resolution mode. The raw image was converted to greyscale, then digitally manipulated to eliminate undesirable shadows, and to make the luminances of the major surfaces more equal. The latter manipulation facilitated a greater range of luminance variations in the experiment. Greyscale was chosen to eliminate chromatic effects from the evaluation of luminance, a common simplification in lighting experiments. The final base image, and the surfaces into which it was divided, is shown in Figure 1. The manipulated surfaces occupied 72% of the total image area. The main surfaces that were not manipulated in the experiment were the computer and the partition-mounted storage element to the right.

The Software

The genetic algorithm—This experiment followed the general model from Johnston's work, in which an analogy between Darwinian natural selection and the selection of the most attractive image is the basis of the technique. To begin the analogy of genetic survival of the fit, we need to define a "gene" for the image of a lit office space. In this case we have six luminous surfaces, as shown in Figure 1.
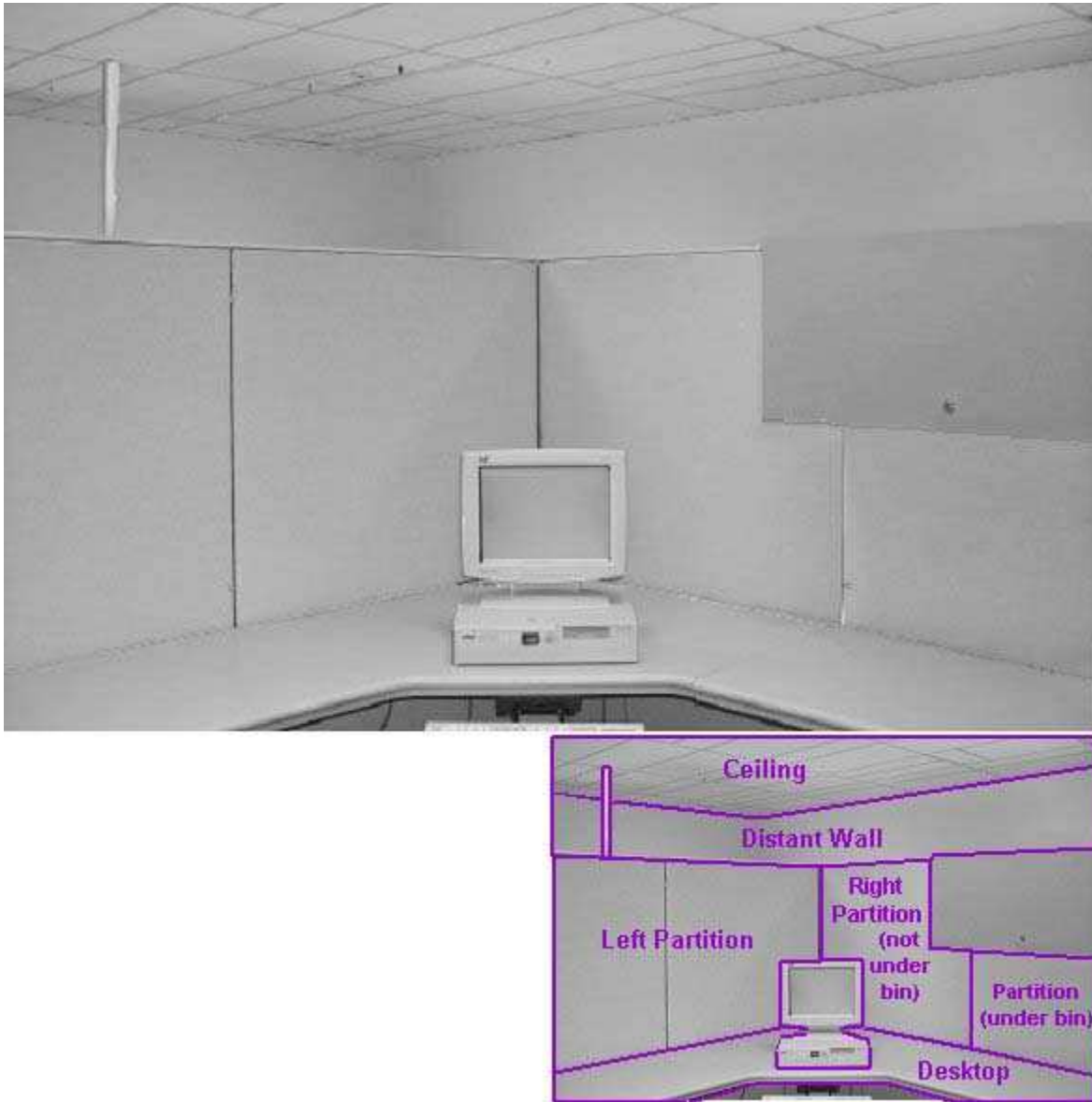

Figure 1 – The base image used in the experiment, and the six surfaces that were independently manipulated in luminance.

We modify the luminance of each surface by changing the greyscale value in the digital image. There were 32 possible levels of luminance for each surface, arranged on an almost linear scale of

increments. Each increment was equivalent to a change in 5 greylevels on the 256 level (8-bit) monochrome scale. Expressed in binary terms, the luminance of each surface varied between 00000 and 11111 (between 0 and 31 in decimal terms, which corresponded to a greyscale range of around 70 to 230, and a luminance range of approximately 3 cd/m^2 to 110 cd/m^2, as presented to the participant). These binary expressions form a "luminance gene sequence" for the surface. So, for example, a luminance level of 9 for a surface, would be represented by the gene 01001, a level of 22 as 10110, and so on. The same can be done for each of the six surfaces, resulting in a 30-digit binary string, or in genetic language, a "phenotype", which uniquely represents the combination of luminances in a particular image. Figure 2 illustrates two examples.

Left Partition	Right Partition (not under bin)	Partition (under bin)	Desktop	Distant Wall	Ceiling
01001	01001	10110	00101	10000	11001
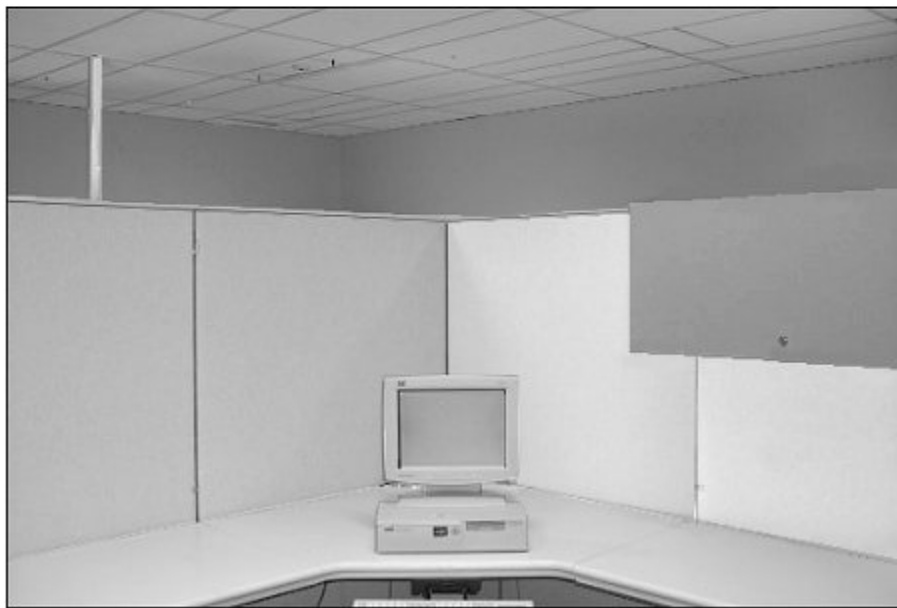
Left Partition	Right Partition (not under bin)	Partition (under bin)	Desktop	Distant Wall	Ceiling
10011	11001	11100	10110	01000	01100

Figure 2 – Two example combination of surface luminances. Below each is the binary phenotype that represents the combinations of luminances, made up of 5-digit genes for each surface.

In the animal kingdom, according the Darwinian evolution, it is the fit individuals that survive, have most offspring, and thus most influence the next generation. Applying a genetic algorithm in our experiment means that the most attractive (our measure of fitness) images are the ones that influence the next generation of images. Two of the most attractive images from a "population of images" are selected as "parent" images, and "reproduce" to create "child" images. Just as in the animal kingdom, these children are the product of their parents' genes and display similarity to their parents' features. We mimic sexual reproduction with operations on the binary strings called crossover and mutation (see Figure 3).
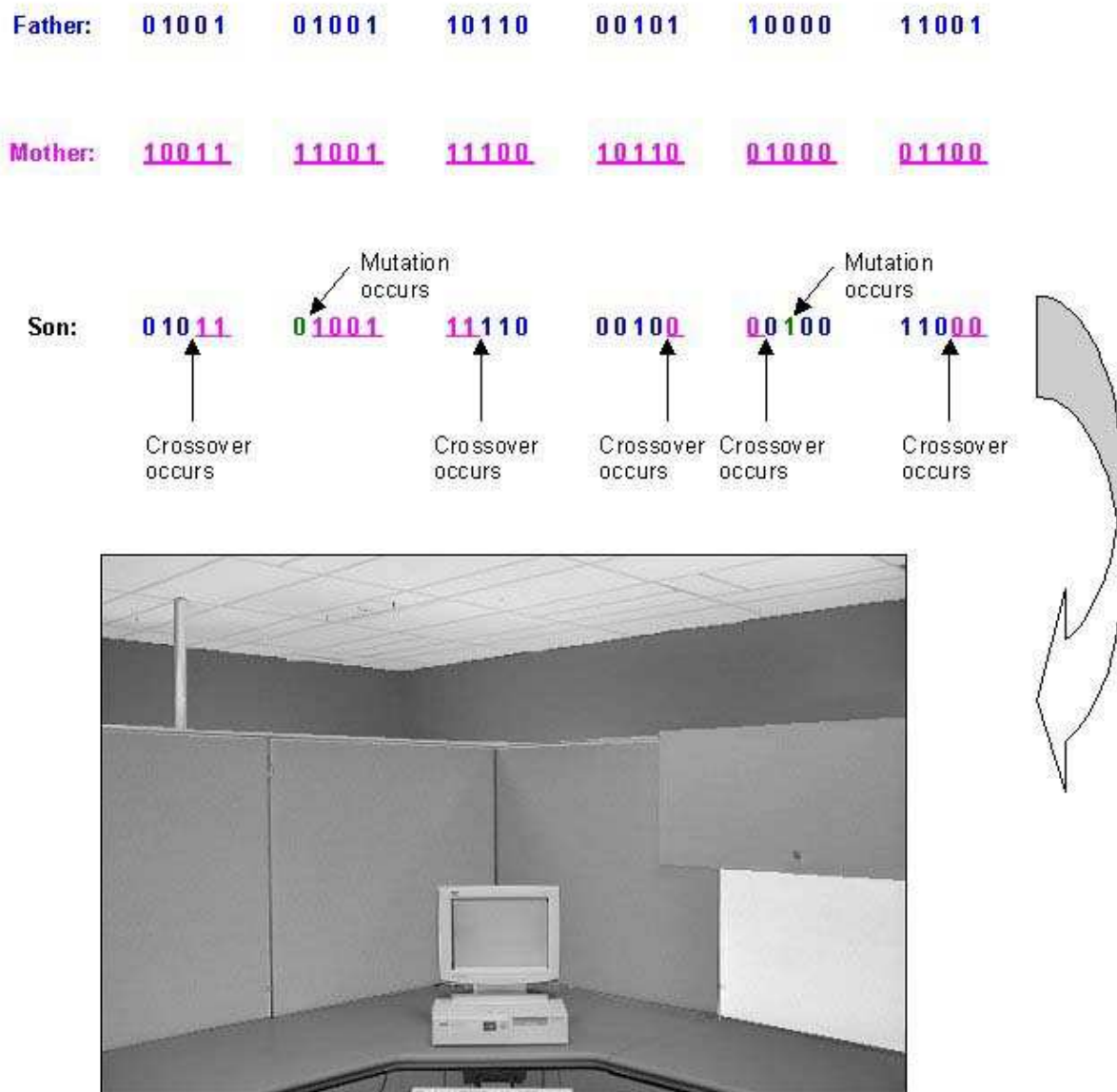
Figure 3 – Crossover and mutation from parents' phenotypes (see Figure 2) to create a son phenotype; mother's genes are underlined. Resulting combination of surface luminances is shown.

In our version, each set of two parents produces two offspring; to follow the analogy, we call the two parents the "father" and the "mother", and the two children the "son" and the "daughter". To create the son's phenotype, we start with the first binary digit in the father's phenotype. For each digit,

reading from left to right, we randomly test to see if crossover occurs, if it does not, the son's digit is a copy of his father's and the next digit of the father's phenotype is tested for crossover. However, if crossover does occur the son's digit is a copy of his mother's and the next digit of the mother's phenotype is tested for crossover. In our experiment, the possibility of crossover at each digit was set at 25%.

As in nature, there is always the (small) possibility of random mutation, which can be very helpful in bringing in new gene combinations which otherwise would not occur. For example, if both the father and the mother have a "0" as the 23rd digit (as in Figure 3), their offspring could never have a "1" as the 23rd digit without random mutation. Random mutations usually produce less fit individuals, and less attractive luminous scenes, but can occasionally get a gene line out of an "evolutionary dead-end". Each of the son's digits is tested for a random mutation; in our experiment, the possibility of random mutation at each digit was set at 4%. The creation of the daughter's phenotype was created in exactly the same way as the son's, except the process begins with the mother's phenotype.

These two children are then presented to the participant and rated for attractiveness. If they are rated more highly than the lowest rated members of the existing population of images then they replace these lower rated members. The process of parent selection and creation of children continues, and the population, on average, becomes more attractive and more homogeneous. Genetic algorithms have proven to be a much more efficient way of finding optimal solutions when the number of variables and possible combinations is large, than simply searching the set of solutions by trial and error (e.g. Johnston[6], p.47).

The software used— Software was written in Visual Basic to present the images, conduct the manipulation of surface luminances according to the genetic algorithm, to administer questionnaires, and to store data. A flow diagram for the software is shown in Figure 4.
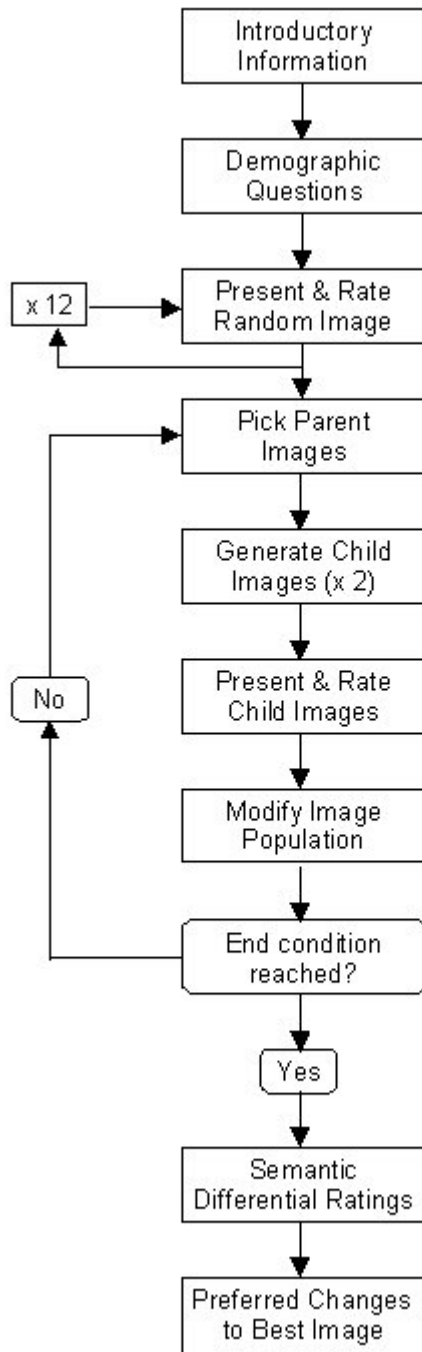
Figure 4 – Overall flow diagram of software used in experiment.

After reading some introductory information, participants were first presented with a basic set of

demographic questions on: Sex, Age, Vision Correction, and Occupation. Following this, the main

task of the session began, the rating of office images. The participants saw an initial set of 12 images, and were asked to rate each one for attractiveness, on a scale of 1 (least attractive possible) to 10 (most attractive possible); there was no limit on the time allowed for this rating. As in the entire session, images were presented one at a time, separated by at least 7.5 seconds during which no image was present on the screen. Each of this initial set of images was simply a random combination of surface luminances. This set of 12 formed the initial population of images.

Then the genetic algorithm process began. Parent images were selected from the population. One parent was always the image with the highest rating. The second parent was selected randomly from the population, but the chance of being selected was weighted according to the image's attractiveness rating. The child images were then presented and rated. After rating each child, an extra element was introduced to help the participant guide the genetic process – they were able to indicate for each surface whether they would prefer it brighter, the same, or darker (see interface in Figure 5).

Figure 5 – Interface for the experimental task. Participants rated the image for overall attractiveness. They then used the boxes at the bottom to indicate their brightness preference for each surface.

The next child was then created according to the genetic process described above, but it was also tested against the participant's preference for each surface. If the child did not meet the set of preferences it was rejected and another child created and tested. The process continued until a satisfactory child was created. Nevertheless, it was not always possible, particularly as the images converged on the participant's optimal image, to meet the participant's preferences for all surfaces. In such cases the criteria for the number of surfaces that met the preferences was decreased from six to five to four and so on until a suitable child was created. The approach of guiding the genetic process was adopted by Johnston and shown to greatly increase the efficiency of the genetic algorithm[17].

The process of parent selection and child creation continued until an end condition was reached. There were three conditions that ended the process:

- An image was given a rating of 10 (this could happen for one of the initial 12 images); or,

- A participant expressed a preference for the same brightness for all six surfaces; or,

- Neither of the two children created from a set of parents was rated more highly than the least attractive member of the prevailing population.

Following the end condition, the participants were asked to rate the appearance of three images on a series of semantic differential (adjective pair) scales. The three images that were rated were the image with the highest rating in the final population (Best image), the image rated third highest in the initial population (Comparison image), and the image with the six surface luminances at the middle of the scale (Neutral image, the same image as that shown in Figure 1). We wanted to test whether the genetic algorithm had in fact led to a preferred image that was significantly different from a non-optimal image. The Comparison was picked from the initial random population so that its creation had not been influenced by the guided genetic algorithm. Nevertheless, we did not want to bias the

evaluations excessively by choosing a Comparison image that had a very low rating. The Best and Comparison images were different for each participant. The Neutral image was chosen because it was the same for all participants, and because it represents the average of all possible images, and is therefore an obvious baseline for evaluations.

For the appearance ratings, we chose 15 bipolar adjective pairs (bright – dim, uniform – non-uniform, interesting – monotonous, pleasant – unpleasant, comfortable – uncomfortable, stimulating – subdued, radiant – gloomy, tense – relaxing, dramatic – diffuse, spacious – cramped, glaring – not-glaring, friendly – hostile, simple – complex, formal – casual, realistic – unrealistic). These adjective pairs were derived from previous research (Hendrick et al. 1977; Eissa & Mahdavi, 2001; Loe et al., 1994; Veitch & Newsham, 1998). The three images were presented in random order with each adjective pair presented one at a time next to the image. The participants gave their rating by moving a cursor on a continuous scale between the two adjectives; according to their strength of feeling; the value recorded ranged from 0 to 100.

Finally, the Best image was recalled to the screen and the participant was asked to indicate, for each surface in turn, whether they would prefer it to be *A lot Brighter, a little Brighter, No Change, a little Darker, A lot Darker*. The image did not change in response to this input. These ratings were another way of assessing how optimal the Best image was.

The Experimental Space

The experiment was conducted using images projected onto a viewing screen using an InFocus model LP530 data projector (see Figure 6 for diagram and photograph).
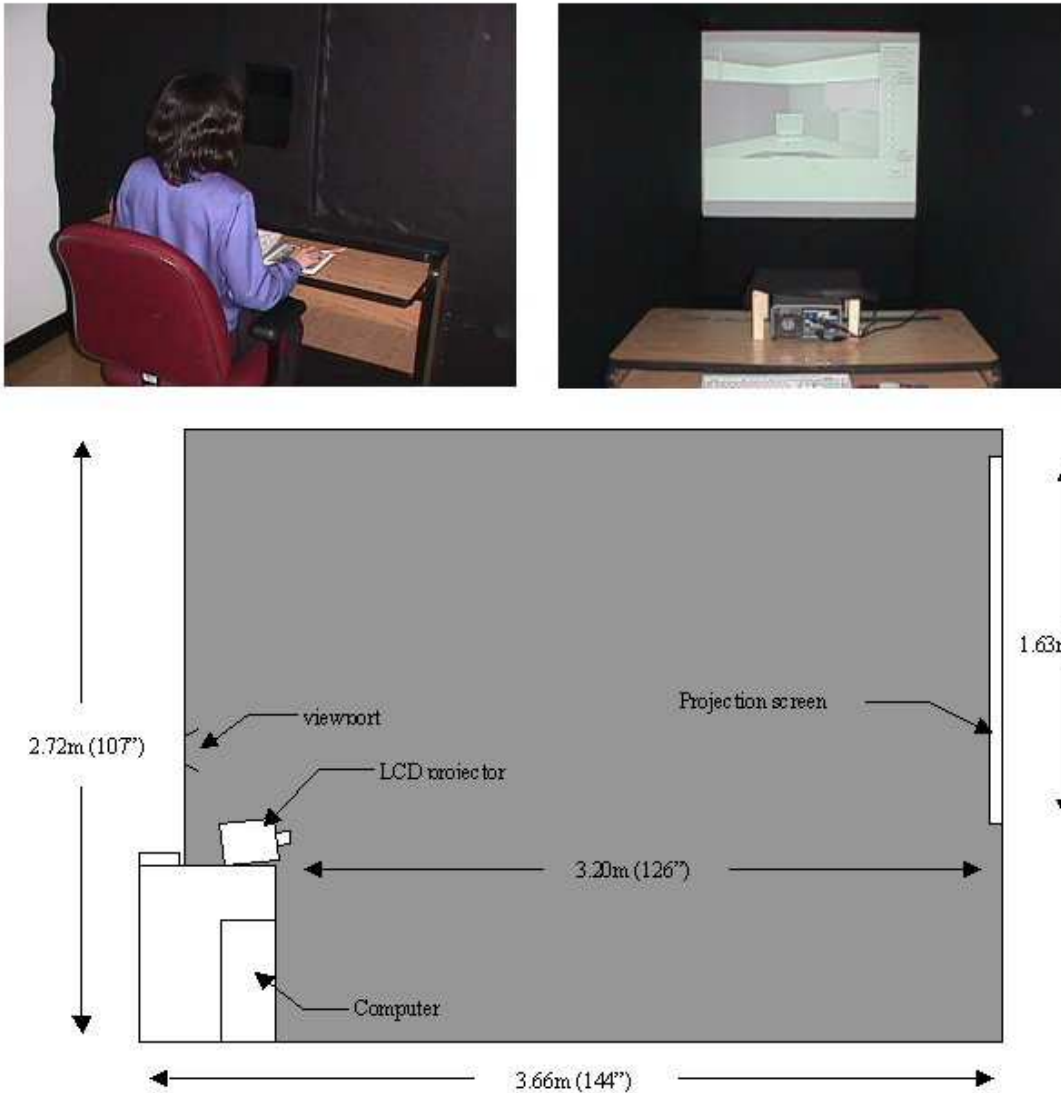
Figure 6 – Experimental set-up. Participants viewed the projected image through a viewport (photo on left). The space into which they looked was black except for the projected image (photo on right, with front wall removed). Diagram shows side elevation, approximately to scale.

Participants sat in a chair and viewed the image through a rectangular slot that could be adjusted in height. The inside of the space into which they looked was completely black except for the image, so

there was no distracting surround. The participant had access to a keyboard and a mouse for answering questionnaires and making ratings.

The distance from the projector to the viewing screen affected both the size of the image and the maximum brightness of the surfaces. We wanted to make both as realistic as possible, but we were limited by the capabilities of the projector. Our final choice gave an image that was 1.52m (60") wide and 1.02m (41") high, such that the computer monitor in the image was 0.20m (8") corner to corner, or about 54% of full size. This allowed us to created surface luminances up to around 110 cd/m^2, typical for a non-daylit office scene, and allowing for a wide range of realistic preferences. The luminance of the computer monitor in the image, which was constant for all images presented, was set to 50 cd/m^2, typical of the luminance we measured for light grey on a real monitor; ANSI/IESNA RP-1[18] suggests typical screen luminances of 50 cd/m^2.

Physical Measurements

Prior to beginning data collection we used a Topcon BM3 luminance meter to provide a calibration between grey level and luminance on the viewing screen. This calibration is shown in Figure 7.

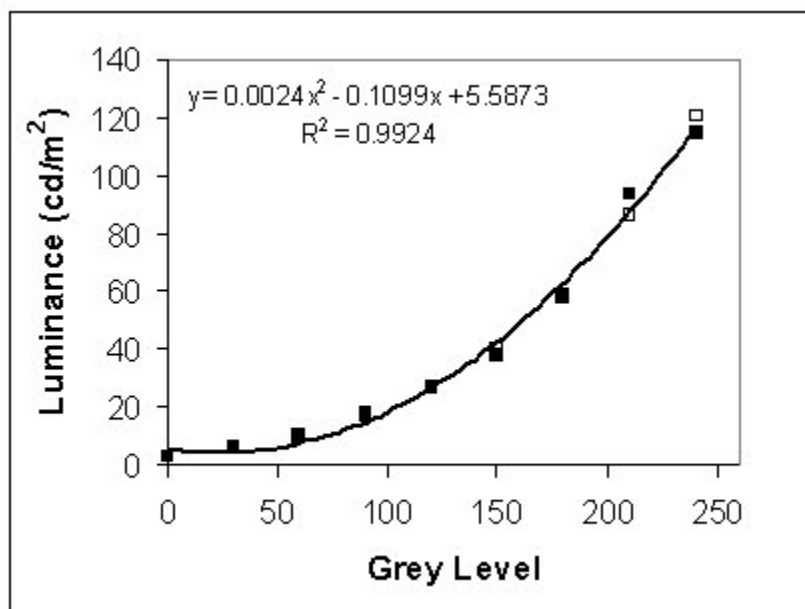$$y = 0.0024x^2 - 0.1099x + 5.5873$$
$$R^2 = 0.9924$$

Figure 7 – Calibration of projector screen luminance vs. image pixel grey level. Open squares show data from Feb. 2002, closed squares show data from June 2002.

The calibration was maintained across two data collection periods four months apart. However, we did notice that from day-to-day the projector was not always consistent in the luminance produced. This repetition error was only around 2 cd/m^2 at 50 cd/m^2, but rose to around 10 cd/m^2 at 110 cd/m^2.

Participants

The 40 participants (23 male, 17 female) were staff from our own institute, and were recruited by a general e-mail announcement. We wanted participants who were naive with respect to lighting, and therefore participants were drawn from all groups within the institute except those involved with indoor environment issues. Participation was voluntary, and the reward for participation was limited to a free coffee and snack. Data was collected in two phases in February and June, 2002. Once initial data collection in each phase was complete, a total of 10 participants were chosen at random and asked to repeat the experiment. Participant characteristics, from their self-reported demographic data, are shown in Table 1.

Experimental Procedure

After arriving at the experimental site (which was set up in a building close to the participants' own offices), participants were given a general description of the experiment both verbally and on paper, and asked to sign a consent form. All further information, instructions and tasks were presented on screen, though the experimenter remained close by to answer any arising questions. The entire procedure was approved by our organisation's Research Ethics Board.

Completion of the on-screen part of the experimental procedure took a mean time of 21:13 (min:sec); s.d. = 7:53 for those doing the experiment for the first time (n=40). Repeaters were somewhat faster, taking a mean time of 15:06 (s.d. = 2:16).

Results

We present the results of the experiment as they pertain to specific questions we wanted to address.

Did the Genetic Algorithm Lead to a Highly-Rated Image?

There were two direct measures pertaining to this question: the attractiveness ratings of the Best image (Figure 8), and the preferred changes in the Best image expressed at the end of the session (Figure 9).
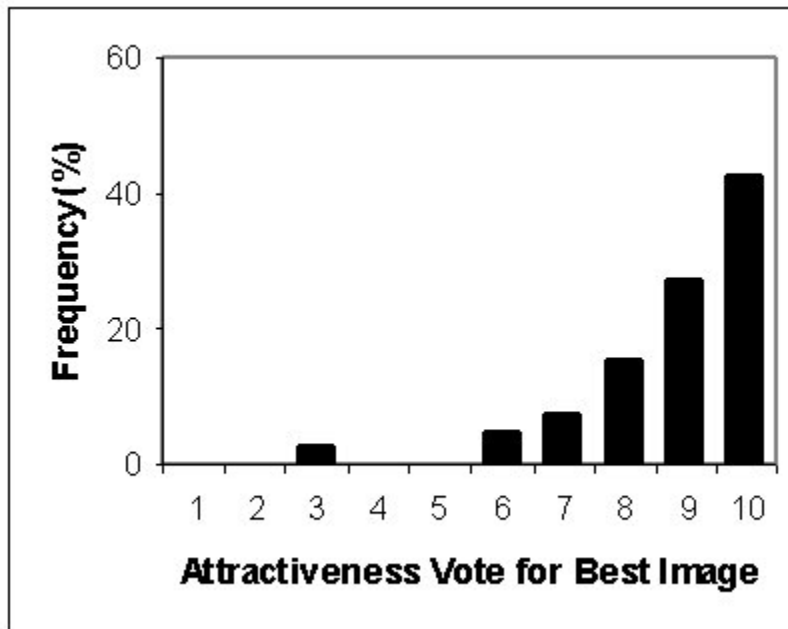
Figure 8 – Histogram of the attractiveness rating of the Best image for participants doing the experiment for the first time (n=40).
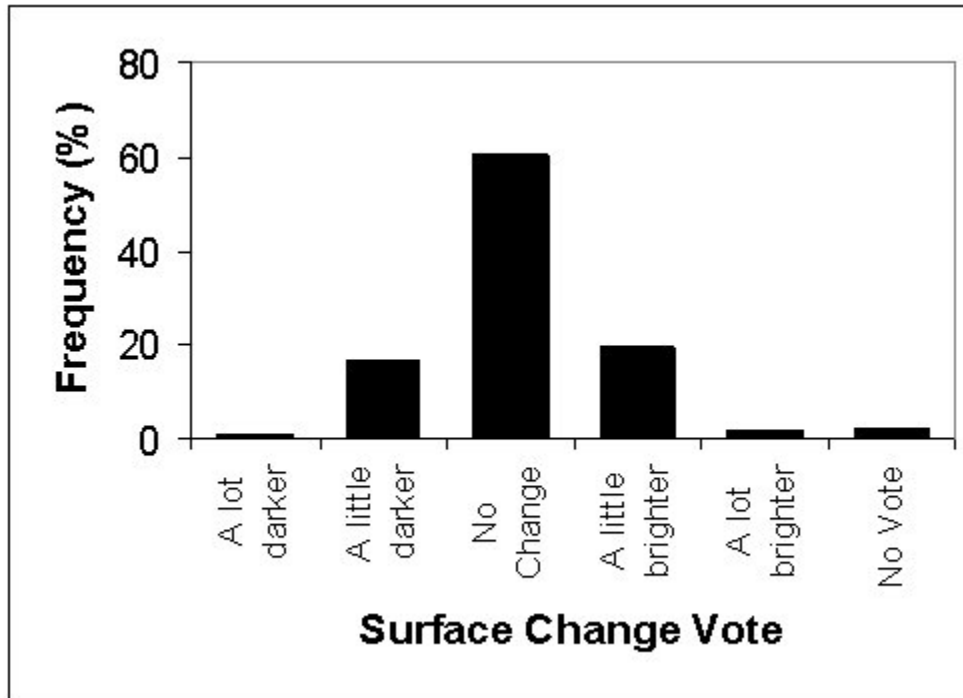
Figure 9 – Histogram of the preference for change in brightness of the six surfaces in the image, for the Best image, for participants doing the experiment for the first time. Data for all surfaces are grouped (n=235).

The mean attractiveness rating of the Best image (scale of 1 to 10) was 8.8 (s.d. = 1.5), and the modal rating was 10. When offered the opportunity to express a final preference for a change in surface brightnesses, 60% of the votes were for no change, and 97% for 'No Change' or 'A little darker' or 'A little brighter'.

A measure of the efficiency of the algorithm in reaching the Best image is the number of images seen by the participants. No participant gave an attractiveness rating of 10 to any of the initial 12 random images; all progressed to the stage where images were generated according to the genetic algorithm. The mean number of images seen was 22.9 (s.d. = 6.0).

Although the Best image was not perfectly optimal (mean attractiveness rating was 8.8, not 10), it was rated very highly, and preferences for change were small. Also, remember that participants were viewing a simple, grey image and some participants might not have felt able to give a rating of 10 no matter what the combination of luminances. The Best image was achieved after viewing relatively few images. To summarize, the genetic algorithm was quite successful. Modifications to the algorithm, the magnitude of brightness changes per level, and the probabilities of mutation and crossover, might improve the algorithm still further in a future study, and working with a more interesting image might be more discriminating.

Are the Most Attractive (Best) Images Rated Differently than a Non-Optimal Images?

For the method to be useful, it must produce optimal images that are rated differently compared to non-optimal images, and in the expected manner – that is to say, observers must be able to reliably discriminate between preferred and non-preferred images. One way we explored this was through the semantic differential appearance ratings of the Best, Comparison and Neutral images. Remember, both the Best and Comparison images differed from participant to participant and had been seen once prior to semantic differential rating, whereas the Neutral image was the same for all participants, and had not been seen before. Figure 10 presents an example Best and Comparison image for one of the participants; the Neutral image is shown in Figure 1.

Figure 10 - Example from one participant of a Best image on the left (in this case, given an attractiveness rating of 10) and its Comparison image on the right (which received an attractiveness rating of 4).

Figure 11 shows the mean ratings for the Best, Comparison and Neutral images for each of the adjective pairs, for participants doing the experiment for the first time (n=40).
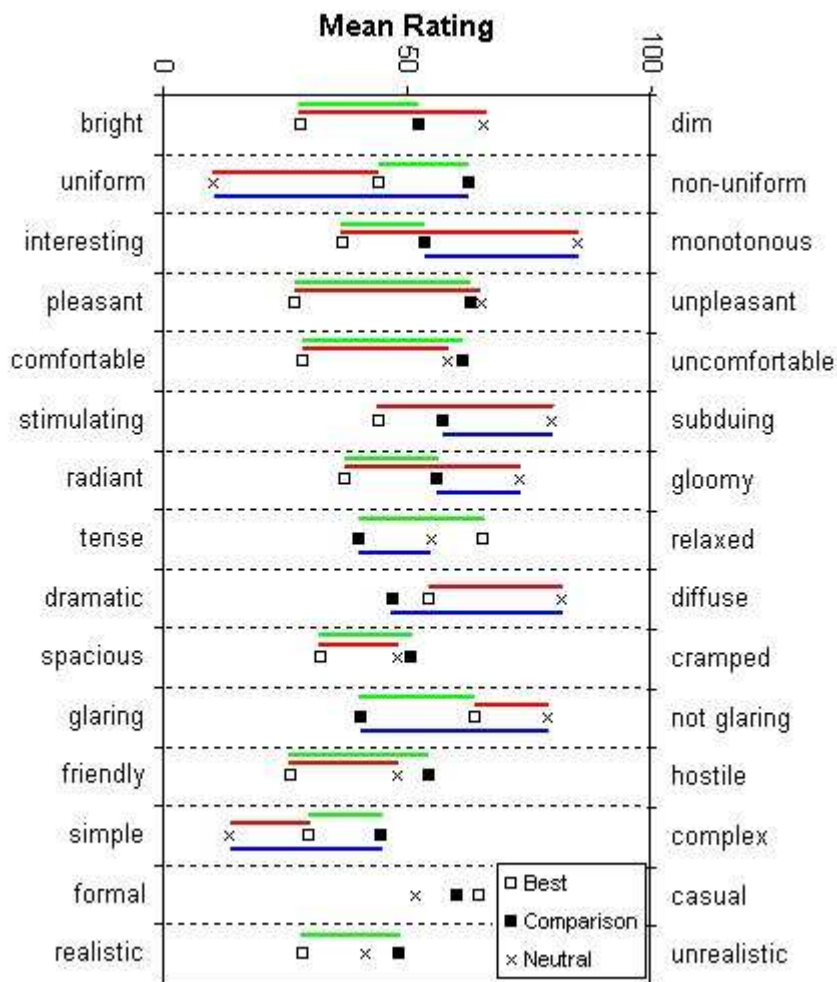
Figure 11 – Comparison of mean semantic differential appearance ratings for the Best, Comparison and Neutral images, for participants doing the experiment for the first time (n=40). A line between two symbols indicates a significant univariate difference (p <0.0167) between those two image types for that adjective pair.

We conducted statistical analyses to check for differences between the ratings. There were *a priori* reasons to compare all image pairs (Best vs. Comparison, Best vs. Neutral, Neutral vs. Comparison). Nevertheless, we divided the usual alpha of 0.05 by 3 for all tests of significance to reduce the possibility of Type I statistical errors; therefore, our critical alpha level was .0167 for all significance tests. We examined within-subjects planned comparisons between images (Best vs. Comparison, Best vs. Neutral, or Neutral vs. Comparison). We first conducted overall multivariate analyses of variance (MANOVA) for the three planned comparisons to check for an overall difference across all adjective pairs. It is usual practice (to control for Type I statistical errors) to test for univariate differences (difference on single adjective pairs) only if the overall MANOVA is significant. All three MANOVAs were statistically significant, and we are fully justified in reporting the univariate tests. The results of the statistical tests are shown in Table 2.

The Best image is significantly more pleasant, more comfortable, and more friendly than both the Comparison and Neutral images. This is expected because the Best image was selected through the genetic algorithm process on the basis of its attractiveness. Other significant univariate effects offer clues as to what properties of the Best image make it more attractive, on average. The Best image is also rated as significantly brighter, less glaring, and more spacious than both the Comparison and Neutral images. The Best image is also rated significantly more uniform and simple than the Comparison image, but significantly less uniform and simple than the Neutral image. In other words, some degree of uniformity is a positive attribute, but excessive uniformity is viewed negatively.

Are Images Perceived in the Same Way as Real Spaces?

Again, for this method to have value, it is important that participants interpret the patterns of light and dark in the viewed images as the lighting of a real scene. We explored this with a factor analysis of the semantic differential appearance ratings. Factor analysis techniques find correlations between

many scales, based on the supposition that responses to the individual scales represent facets of a response to a smaller number of higher-level concepts. We expected that if the participants interpreted the images as the lighting of a real scene then the factor structure (the way in which individual scales load on factors) would be the same as the factor structure derived from ratings of a real scene irrespective of the absolute ratings on the individual scales. The appearance ratings from our experiment were subjected to a factor analysis (principal components). The ratings of the Best, Comparison and Neutral images were analysed independently, and the results are shown in Table 3.

Also shown in Table 3 is the factor analysis of Veitch and Newsham[15] data. The image we used for this experiment was taken in a mock-up office space that had been previously used for a human factors study of lighting quality (Veitch & Newsham[15]). In that study nearly 300 participants experienced nine different office lighting designs (between-subjects) for a day, doing simulated office tasks and completing a variety of questionnaires. The participants completed a similar set of semantic differential appearance ratings on which factor analysis was performed. The ratings in Veitch and Newsham included more adjective pairs than we used in this study, only shared items are shown in the table, except for two adjective pairs, 'friendly-hostile' and 'realistic-unrealistic', which were used in the present study and not used in Veitch and Newsham. In all analyses, using a Eigenvalue of 1 as a criterion, a four-factor solution was suggested. We used a conservative criterion for factor loadings, highlighting only those unrotated loadings greater than 0.5.

The factor structure from Veitch and Newsham, data collected from people who had occupied a real space for several hours, is very simple and interpretable. Most scales loaded on a single factor, except for 'glaring – not glaring', which did not load at all, and 'formal – casual', which loaded on two factors. More scales loaded on Factor 1 than any other, and this Factor was interpreted as the general concept of image attractiveness. Factor 2 comprised two scales both concerned with image variability or uniformity. Factor 3 comprised only one scale, but addressed the important issue of image brightness.

In comparison, the factor structure of the Best, Comparison and Neutral images from the present study is more complex. The Neutral image has a very similar set of loadings on Factor 1, but not on

the other Factors. The Best image structure is the only one that reproduces the same Factors 2 and

3 seen in the Veitch and Newsham data, and has some similarity on Factor 1. This provides some

encouragement that the luminous environment in Best image, at least, is being perceived in a similar

way as a similar real space. It is not surprising that the Best image is, in this manner, being

perceived as more realistic than the Comparison image, given the nature of the random distribution of

brightnesses in the Comparison image. Note also that in Figure 11, for the univariate test on the

'realistic – unrealistic' scale, the Best image is rated as significantly more realistic than the

Comparison image.

Are Preferred Luminances and Ratios the same as Those Derived from Experiments in Real Spaces?

Table 4 presents a summary of the luminance information from the 40 Best images for those

participants doing the experiment for the first time. Data is presented for the six surfaces that were

independently manipulated, for meaningful combinations of surfaces, and for important ratios.

The individual median surface luminances in the set of Best images is shown pictorially in Figure 12[@].

Figure 12 – (Approx.) Average preferred luminance.

[@] The translation to the image is only approximate because the numerical average does not necessarily correspond with an integer gene value.

It is interesting to compare this with the base (Neutral) image in Figure 1 in which each surface is in the middle of the range of possible values. The most obvious difference is in the brightness of the ceiling. The importance of ceiling brightness as driver of image preference will be explored later in this paper.

To check further the validity of this method we compared the preferred luminous conditions in the Best images to those derived from human factors experiments in real spaces. All of the studies we refer to in this section excluded daylight and derived preferences for electric lighting only, analogous with our study. Veitch and Newsham[16] conducted a study in the same laboratory space in which the photograph used in this study was taken. Participants had dimmable control over three lighting circuits (one indirect and two direct) as well as on-off control over an undershelf task light. Participants occupied the space for an 8-hour day during which they conducted typical office tasks, predominantly computer-based. Half the sample of 94 had control at the start of the day, the other half at the end of the day. One value reported was the mean luminance in an area that included the partitions behind the computer and under a binder bin, part of the desktop and part of the binder bin – this area was chosen to represent the 40° horizontal band of field of view from Loe et al.[19]. The median luminance chosen by participants in this area was 39.2 cd/m^2 (min. = 11.5, max. = 61.0). The nearest equivalent value from this study is probably the mean of ABCD (all the partitions and desktop taken together, median = 47.8 cd/m^2 (min. = 29.4, max. = 91.7). Note that the value from Veitch and Newsham included part of the binder bin – a very dark area (< 10 cd/m^2) occupying around $^1/_{12}$ of the field of view, which was not included in the mean of ABCD. Removing the binder bin from the Veitch and Newsham area would increase the median luminance by around 10%, making the comparison to this study even better.

Berrutto et al.[20] also gave participants dimming control over a variety of luminaires in small private offices. Exposures were limited to 20 minutes, and data were collected separately for different tasks. They concluded that for non-VDT tasks wall luminance at eye level should be around 60-65 cd/m^2, and for VDT tasks the luminances around the screen should be equal or lower than the luminance of

the screen. In our study, the screen luminance was 50 cd/m^2 and the immediate surround luminance, perhaps best expressed by the average of AB, had a median of 50.4 cd/m^2.

Van Ooyen et al.[21] presented participants with different office luminous environments by manipulating light distributions and changing the reflectivity of surfaces; working plane illuminance was maintained at around 750 lux. The spaces were private two-person offices, and data were collected separately for different tasks. For non-VDT tasks, preferred wall luminances were 30 to 60 cd/m^2, and preferred working plane luminances were 45 to 105 cd/m^2. For VDT work the values were reduced: preferred wall luminances were 20 to 45 cd/m^2, and preferred working plane luminances were 40 to 65 cd/m^2. Van Ooyen et al. reported that the preferred ratio of working plane luminance to wall luminance was 1.33, our equivalent ratio D:ABC was 0.77. ANSI/IESNA RP-1[18] suggests no favoured direction for recommendations for adjacent surface luminance ratios in offices. In this context, it is interesting to note that the reciprocal of 0.77 is 1.30, very close to the preferred ratio reported by Van Ooyen et al.

Loe et al.[19] had observers rate a small conference room from a point of view equivalent to the room's entrance. Lighting conditions were manipulated by the experimenters using a variety of luminaires. They concluded that for 'visual lightness' the preferred average luminance in a horizontal band of field of view 40° wide should be >= 30 cd/m^2. Again note that the minimum value for the mean of ABCD in our study was 29.4 cd/m^2, and a value less than 30 cd/m^2 was present in the Best image of only one participant.

It is interesting to note that the median preferred ratio of the luminance of the adjacent partition walls (A:B) was 0.96, very close to 1. This is as predicted, as any real ambient lighting system would produce approximately equal luminances on such adjacent surfaces that are free of obstructions.

A final useful comparison to data from real spaces lies not in the comparison of averages, but in the comparison of variability in individual preference. We observed a wide variety of preferred luminances in the Best images of our participants. This is encouraging because studies of individual preference in real spaces also report a wide variety of choices. We performed a quantitative comparison to the results of Veitch and Newsham[16]. In Figure 13 we plot two measures of the

frequency of preferred brightness in the field of view, one from this study and one from Veitch and Newsham[16].
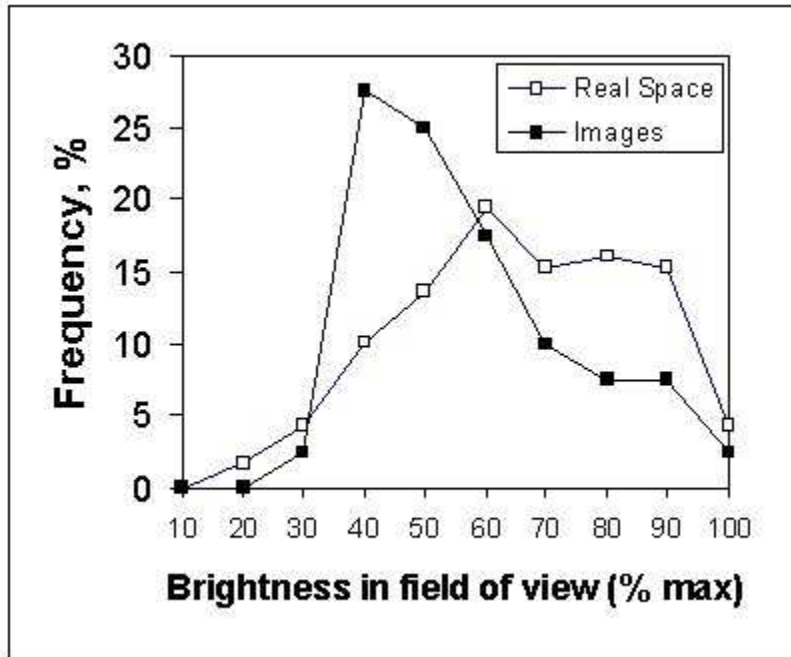
Figure 13 – Preferred brightness in the field of view, as a percentage of the maximum available, for this study and a study done in a similar real space (Veitch & Newsham).

To facilitate comparison between the different measures we have expressed preference in terms of the (approximate) maximum available choice in each respective experiment. For this study we plot the weighted average of ABCD (see Table 4). For Veitch and Newsham we plot the mean luminance in the 40° horizontal band of field of view (as described above). Although the two curves are skewed differently, they illustrate similar forms of variability in individual preference.

Taken together, these comparisons show that the preferred luminances derived from our study compare well with the preferred luminances from other studies using more traditional methods, and with expectations for realistic lighting. As such, these comparisons reinforce the hypothesis that the images are perceived in the same way as real spaces.

Do Subjective Ratings Correlate with Photometric Descriptors?

This is a question fundamental to lighting quality research. If occupant ratings of luminous environments, particularly ratings of attractiveness, could be reliably correlated with photometric descriptors that are easily calculated prior to a space's construction, this would be a major step forward for lighting design. Unfortunately, this has proven difficult, though some progress has been made[15,19,22,23].

From the semantic differential ratings we concluded that our participants were seeking attractive spaces that were bright but not glaring, and that had a degree of uniformity without being too uniform. From the preferred luminances we observed that a bright ceiling was a popular choice of our participants. The obvious next step is to examine whether ceiling luminance, or other luminance measures, are predictive of attractiveness, brightness, and uniformity.

Rather than use all 15 of the semantic differential scales, we reduced the number of subjective outcomes to a smaller set for simplicity. We used the factor analyses from Section 3.3, together with the factor analysis from a prior pilot study[14] to derive three measures from the individual semantic differential scales that have proven most reliable across studies and image types. Our composite measure of attractiveness was the mean of the ratings on the following individual scales: 'friendly – hostile', 'pleasant – unpleasant', 'comfortable – uncomfortable', 'spacious – cramped', and 'tense – relaxed' (reverse coded). The higher the value of this composite measure the less attractive the image was rated, therefore we named the measure F_UNATTRAC. Similarly, our composite measure of uniformity was the mean of the ratings on the 'uniform – nonuniform', and 'simple – complex' scales. The higher the value of this composite measure the less uniform the image was rated, therefore we named the measure F_NONUNIFRM. Our measure of brightness was formed from only one individual scale ('bright – dim'). The higher the value of this measure the dimmer the image was rated, therefore we named this measure F_DIM.

We then looked at linear regressions of these three subjective outcomes vs. a small set of photometric values expected to be important. If these relationships are reliable then they should

apply across all images, not just the set of Best images. Therefore we included ratings and photometric values for all three images (Best, Comparison, Neutral) that each participant rated using the semantic differential scales. This gave us 120 data points per regression (3 images x 40 participants doing the experiment for the first time). Because each participant is providing 3 data points (2 of which differ from participant to participant on luminance), the points cannot be considered independent and we cannot perform simple regressions. Not only would it be incorrect to do so, it would be to our disadvantage because we would be ignoring within-subject variance; accounting for within-subject variance increases our explanatory power. The relatively new statistical technique of Hierarchical Linear Modelling (HLM, or mixed regression)[24-25] accounts for the within-subject effects. Conceptually, this analysis consists of creating separate regression lines for each participant, and then testing the distribution of regression weights (slopes and intercepts) against the null hypothesis that the average regression weight equals zero. The technique also produces a single best-fit regression line across all data points, which we will report in this paper[#].

We will start with uniformity. A simple photometric descriptor of luminance nonuniformity is the root mean square of the difference between the six individual surface luminances and the weighted-mean of all six. This parameter is denoted RMSLNUNI, and is expressed mathematically as:

$$RMSLNUNI \quad = \quad \sqrt{(\Sigma_{i=A..F}(L_i - L^*)^2)}$$

where,
L_i = luminance of an individual surface (cd/m^2)
L^* = weighted-mean luminance of all six surfaces (cd/m^2)

The higher value of RMSLUNI the larger the overall differences between each surface and the mean of all surfaces, and the higher the photometric nonuniformity. A "spaghetti" plot of F_NONUNIFRM vs. RMSLNUNI is shown in Figure 14, with separate lines joining the three data points from each participant.

[#] The model we used was a random intercept and random slope model with no centering, with one photometric predictor at level-1, and no level-2 predictors; we were interested in explaining whether photometric variables predicted semantic ratings (level-1), and not in investigating what participant characteristics might have led to differences in ratings between participants (level-2).
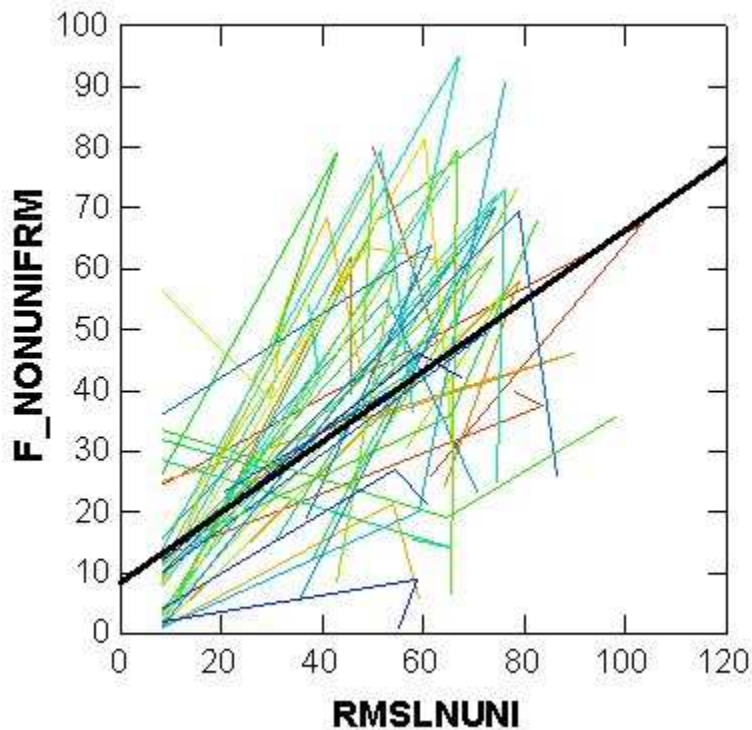
Figure 14 – "Spaghetti" plot of 40 individuals' subjective ratings of uniformity

(F_NONUNIFRM) vs. a photometric descriptor of nonuniformity (RMSLNUNI). The best-fit

linear regression line is superimposed.

Although there is much between-subject variability, there is a general trend in the expected direction.

The HLM analysis indicates that this linear trend is significant: intercept = 8 (t = 2.54, d.f. = 39, $p <$

0.05); slope = 0.64 (t = 9.14, d.f. = 39, $p <$ 0.01). The percentage of variance explained is high (0.52).

The intercept is very close to the origin. This is expected; an image where all six of the manipulated

surfaces had the same luminance (RMSLNUNI = 0) should be subjectively rated as very uniform

(F_NONUNIFRM close to 0)[$].

We then follow a similar process for F_DIM and F_UNATTRAC. The results are summarized in Table

5. In the case of F_DIM the obvious first predictor to try is the mean brightness of the six manipulated

---

[$] We would not expect the intercept to go through F_NONUNIFRM = 0 because the six surfaces do
not comprise the entire image, and the parts of the image not manipulated in the experiment do not
have identical luminances.

surfaces, L*, and indeed, this relationship is significant and in the expected direction (remember, low values of F_DIM mean the image is rated as brighter). The intercept is very close to 100, this is expected, an image where all six of the manipulated surfaces were black (L* ≈ 0) should be subjectively rated as very dim (F_DIM close to 100). However, the luminance of the ceiling (L_F) alone explains almost as much variance as L*. RMSLNUNI is also a significant predictor of F_DIM: images that are photometrically less uniform are subjectively rated as brighter. Tiller & Veitch[26] found a similar result. In their study participants viewed side-by-side, full-scale, mock-up offices under two lighting conditions of differing uniformities, and used a dimmer to match the room brightnesses between the two rooms. They found that the room with the lower uniformity required 5 to 10% less working plane illuminance to match the brightness of the more uniform room.

Whereas there are obvious photometric predictors for ratings of uniformity and brightness, there are no such obvious predictors for ratings of attractiveness. We tried several possibilities and found that the only single surface luminance that was both a significant predictor and explained a substantial proportion of variance (0.18) was ceiling luminance, L_F. There were other more complex photometric measures that were significant predictors, all contained L_F, but they did not explain as much variance as L_F alone[&]. Table 5 shows that the brighter the ceiling, the higher the rating of attractiveness (remember, low values of F_UNATTRAC mean the image is rated as more attractive).

Unfortunately, our number of participants was relatively small, as was the number of observations per participant, which does not allow us to explore multiple predictors using HLM on our data. In future studies, with more observations, use of multiple predictors might provide more explanatory power. Nevertheless, the fact that we can predict, to some extent, subjective evaluations from simple photometric measures is very encouraging.

How Repeatable is the Genetic Algorithm Method?

We addressed this question by comparing the Best images from two trials for the 10 participants who repeated the experiment. However, any findings can only be considered preliminary given the small

---

[&] Other predictors tried were: $L_A \ldots L_F$ individually, L*, L_{AB} (the weighted mean of L_A and L_B), AR_{AB} (=L_A/L_B if $L_A > L_B$, else=L_B/L_A), L_{EF} (the weighted mean of L_E and L_F), L_F/L_{AB}, L_F/AR_{AB}, RMSLUNI, R_{FAB} (=L_F/L_{AB} if $L_F > L_{AB}$, else=L_{AB}/L_F), $\log_{10}(R_{FAB})$.

sample size. It is noteworthy that participants in the second experimental trial took, on average, less time than they took during the first trial (15:06 vs. 21:02) and saw fewer images (18.2 vs. 23.8). This might indicate a more hasty approach on behalf of the participants in the second trial that could potentially compromise the comparison of Best images. The mean attractiveness vote of the Best images from the second trial was lower than from the first trial (7.7 vs. 8.9), though a two-tailed paired t-test was not significant. Neither was there any significant difference between the luminances of the six surfaces in the two sets of Best images, nor between the F_DIM, F_NONUNIFRM, and F_UNATTRAC ratings.

A visual inspection of the Best Images from the repeaters showed that some participants produced almost identical Best images between the two trials, whereas others did not. Nevertheless, in the latter case there are clues that certain key elements of the luminance pattern were retained; e.g. a preference for a relatively bright ceiling. A larger sample of repeaters would be necessary to explore this issue further.

Further Discussion

The results of the present study both support and extend the findings of our previous pilot study[14]. Note that one of the criticisms of Newsham et al.[14] was that the participants were a relatively small group (n=22) of lighting experts, and the present study employed a larger group (n=40) of naïve participants; the similarity of the results might suggest that lighting experts are not that different from the general public after all! In both studies, the genetic algorithm process produced a highly rated optimal image in a relatively short time. The mean attractiveness rating of the Best image in the present study was 8.8, compared to 8.2 in Newsham et al.[14], and the mean number of images rated was 22.9 vs. 22.5, respectively. Identical semantic differential ratings were used in Newsham et al.[14], however, the Best image was compared to the Comparison image (75th percentile image in the initial population) only, there was no Neutral image for comparison. A similar pattern of significant differences between Best and Comparison images occurred in both experiments, with similar effect sizes in most cases. It is noteworthy that although the overall MANOVA was not significant in Newsham et al.[14], it was significant in the present study, with a larger sample size. In both studies we

observed that the Best image was rated as significantly more uniform than the Comparison image. This surprised some readers of Newsham et al.[14], who thought uniformity was not necessarily a desirable trait in office lighting design. The present study clarifies this, showing that a Neutral image is rated as even more uniform than the Best image, but is also rated as less attractive. The conclusion is that some uniformity is a good thing, excessive uniformity (monotony) is not. Both studies reached similar conclusions about the factor structure of the semantic differential ratings. In both studies the preferred luminances are similar to those found by experiment in real spaces, though the preferred luminances in the present study tend to be higher than those in Newsham et al.[14]. The only major difference in preferred luminance ratio is desktop:partitions (D:ABC). In the present study the mean preference was for the partitions to be brighter (ratio=0.77), in the Newsham et al.[14] the preference was for the desktop to be brighter (ratio=1.21). Note that ANSI/IESNA RP-1[18] suggests no favoured direction for recommendations for adjacent surface luminance ratios in offices, and that the reciprocal of 0.77 is 1.30, not dissimilar from 1.21. In both studies, the mean preferred luminance pattern presents a ceiling that is substantially brighter than other surfaces. Newsham et al.14 did not explore the relationships between photometric descriptors and subjective ratings that were successfully addressed in the present study.

Despite this studies' encouraging results, there are potential improvements that could be made to the method in the future. In this study we changed the luminance of objects uniformly across their surface, which emphasized the independence of these changes from any luminaires. In addition, the surface luminances were independent of each other; there were no inter-reflections. Furthermore, no luminaires were visible in the viewed image. This was deliberate; we were interested in participants' fundamental preferences of light and dark without concern as to how this might be achieved in reality. However, there is a danger in this: in reality, surface luminances can be changed via illumination or by changing reflectance, and the two are not necessarily perceptually equivalent (see references in Moeck[13]). Therefore, our approach risks confusing the participant when viewing the scene: are they observing lighting manipulations, reflectance manipulations, or a combination of both? This could be a mechanism for reducing the reality of the experience. Nevertheless, we reiterate that our results compare well with those obtained in real settings. It is also worth noting that natural-looking inter-

reflections could be achieved in our study indirectly. A brighter partition would not directly produce a brighter desktop, but a brighter desktop could be achieved in future generations of the image. All the same, in future work we intend to include realistic luminaires and inter-reflections.

Some participants were frustrated that they could only express their preferences by guiding the evolution process – which did not always give them exactly the result they were expecting – rather than by directly manipulating the brightness of each surface. However, one of the advantages of the genetic algorithm is that it is not deterministic and includes some randomness. This forces participants to consider luminous environments they are not familiar with but which might have advantages, rather than allowing them to immediately select a pattern of brightnesses according to a pre-existing bias. The genetic algorithm allows participants to indicate their preference for a favoured lighting pattern when they see it rather than asking them to "design" lighting from scratch. This is probably most useful for studying the preferences of people who are not lighting experts.

The importance of ceiling luminance indicated by our results is intriguing and worthy of further investigation. Several recent studies performed in real spaces indicate occupant preference for a substantial indirect component in office lighting[16,27,28], which would provide luminance on the ceiling. However, remember that in our study the ceiling luminance was very uniform and the median ceiling luminance in the Best images was 80.5 cd/m^2. A real indirect lighting system with typical fixture spacings would not produce a perfectly uniform ceiling, and would create bright areas on the ceiling directly above the fixtures that would be substantially higher in luminance than 80.5 cd/m^2.

Conclusions

The main findings of this experiment are:

- The genetic algorithm method is quite successful in obtaining a participant's preferred luminance patterns (Best image) in a greyscale image of an office space.
- Participants were sensitive to manipulation of luminances in the image, such that appearance ratings of participants' Best images were significantly different from non-optimal images in the expected directions.

- The preferred surface luminances from the projected images are similar to those from experiments in real settings.

Analysis of the set of Best images chosen by participants in this experiment suggests the following:

- People want spaces that are bright, but not glary, and brightness ratings can be predicted, in part, from overall luminance, ceiling luminance and luminance variability.

- People want spaces that are somewhat uniform, but not monotonous, and uniformity ratings can be predicted, in part, from luminance variability.

- People want attractive spaces, and attractiveness ratings can be predicted, in part, from ceiling luminance.

All of these findings indicate that the genetic algorithm method of deriving preferred patterns of luminance, realised by projecting close to full-scale images at realistic luminances, has promise. If future work reinforces our findings that the results of this method are equivalent in many ways to the results from real settings, this method might be able to replace much more expensive experiments in real settings in some instances.

References

1. Daniel, T.C.; Meitner, M.M. 2000. Representational validity of landscape visualizations: the effects of graphic realism on perceived scenic beauty of forest vistas. *Journal of Environmental Psychology* 21: 61-72.

2. Danford, S.; Willems, E.P. 1975. Subjective responses to architectural displays. *Environment and Behavior* 7(No. 4): 486-516.

3. Rohrmann, B.; Bishop, I. 2002. Subjective responses to computer simulations of urban environments. *Journal of Environmental Psychology* 22: 319-331.

4. Hendrick, C.; Martyniuk, O.; Spencer, T.J.; Flynn, J.E. 1977. Procedures for investigating the effect of light on impression: simulation of a real space by slides. *Environment and Behavior* 9 (No. 4): 491-510.

5. Mahdavi, A.; Eissa, H. 2002. Subjective evaluation of architectural lighting via computationally rendered images. *Journal of the Illuminating Engineering Society* 31(No. 2): 11-20.

6. Johnston, V.S. 1999. *Why We Feel*. Reading: Perseus Books.

7. Johnston, V.S.; Franklin, M. 1993. Is beauty in the eye of the beholder? *Ethology and Sociobiology* 14: 183-199.

8. Ashdown, I. 1994. Non-imaging optics design using genetic algorithms. *Journal of the Illuminating Engineering Society* (Winter): 12-21.

9. Eklund, N.H.; Embrechts, M.J. 2001. Multi-objective optimization of spectra using genetic algorithms. *Journal of the Illuminating Engineering Society* (Summer): 65-72.

10. Chutarat, A.; Norford, L.K. 2001. A new design process using an inverse method: a genetic algorithm for daylighting design. *Proceedings of IESNA Annual Conference (Ottawa)*: 217-229. New York: Illuminating Engineering Society of North America.

11. Kawai, J.K.; Painter, J.S.; Cohen, M.F. 1993. Radioptimization - goal based rendering. *Proceedings of SIGGRAPH (Anaheim)*: 147-154.

12. Schoeneman, C.; Dorsey, J.; Smits, B.; Arvo, J.; Greenberg, D. 1993. Painting with light. *Proceedings of SIGGRAPH (Anaheim)*: 143-146.

13. Moeck, M. 2001. Designed appearance lighting – revisited. *Journal of the Illuminating Engineering Society* (Summer): 53-64.

14. Newsham, G.R.; Marchand, R.G.; Veitch, J.A. 2002. Preferred surface luminances in offices, by evolution: a pilot study. *Proceedings of IESNA Annual Conference (Salt Lake City)*: 375-398. New York: Illuminating Engineering Society of North America.

15. Veitch, J.A.; Newsham, G.R. 1998. Lighting quality and energy-efficiency effects on task performance, mood, health, satisfaction and comfort. *Journal of the Illuminating Engineering Society* (Winter): 107 - 129.

16. Veitch, J.A.; Newsham, G.R. 2000. Preferred luminous conditions in open-plan offices: research and practice recommendations. *Lighting Research and Technology* 32 (No. 4): 199-212.

17. Caldwell, C.; Johnston, V.S. 1991. Tracking a criminal suspect through 'Face-Space' with a genetic algorithm. *Proceedings of the 4th International Conference on Genetic Algorithms (San Diego)*: 416-421. San Mateo: Morgan Kaufmann.

18. ANSI/IESNA. 1993. RP-1, *Office Lighting*, American National Standard, New York: Illuminating Engineering Society of North America.

19. Loe, D. L.; Mansfield, K. P.; Rowlands, E. 1994. Appearance of lit environment and its relevance in lighting design: experimental study. *Lighting Research and Technology* 26(No. 3): 119 – 133.

20. Berrutto, V.; Fontoynont, M.; Avouac-Bastie, P. 1997. Importance of wall luminance on users satisfaction: pilot study on 73 office workers. *Proceedings of Lux Europa – 8th European Lighting Conference (Amsterdam)*: 82-101. Arnhem: NSVV.

21. Van Ooyen, M.H.F.; van de Weijgert, J.A.C; Begemann, S.H.A. 1987. Preferred luminances in offices. *Journal of the Illuminating Engineering Society* (Summer): 152 - 156.

22. Flynn, J.E.; Hendrick, C.; Spencer, T.; Martyniuk. 1979. A guide to methodology procedures for measuring subjective impressions in lighting. *Journal of the Illuminating Engineering Society* 8 (No. 2): 95 - 110.

23. Newsham, G.R.; Veitch, J.A. 2001. Lighting quality recommendations for VDT offices: a new method of derivation. *Lighting Research and Technology* 33 (No. 2): 97-116.

24. Bryk, A.S.; Raudenbush, S.W. 1992. *Hierarchical Linear Models*, Newbury Park: Sage Publications.

25. Hox, J.J. 1995. *Applied Multilevel Analysis*. Amsterdam: TT-Publikaties.

26. Tiller, D.K.; Veitch, J.A. 2000. Perceived room brightness: pilot study on the effect of luminance distribution. *Lighting Research and Technology* 27 (No. 2): 93-101.

27. Houser, K.W.; Tiller, D.K.; Bernecker, C.A.; Mistrick, R.G. 2002. The subjective response to linear fluorescent direct/indirect lighting systems. *Lighting Research and Technology* 34(No. 3): 243 – 264.

28. Govén, T.; Bångens, L.; Persson, B. 2002. Preferred luminance distribution in working areas. *Proceedings of Right Light 5, 5th European Conference on Energy-Efficient Lighting (Nice)*: 87-92. Stockholm: Borg & Co.

Table 1. Participant characteristics.

	Total responses						
Sex		*Female*			*Male*		
initial group	40	17			23		
repeaters	10	5			5		
Age		*18-29*	*30-39*	*40-49*	*50-59*	*60-69*	
initial group	40	1	13	15	10	1	
repeaters	10	0	3	5	2	0	
Correction Lenses		*None*	*Reading Glasses*	*Distance Glasses*	*Bi- or Trifocal Lenses*	*Gradual or Multifocal Lenses*	*Contact Lenses*
initial group	40	11	7	11	5	3	3
repeaters	10	3	2	1	3	0	1
Principal Occupation		*Administrative*		*Technical*		*Professional*	*Managerial*
initial group	40	10		8		16	6
repeaters	10	3		3		3	1

Table 2. Result of MANOVA and univariate effects on appearance ratings for participants doing the experiment for the first time (n=40). Statistical tests were within-subjects on a single independent variable: image type; with two levels. Only statistically significant effects are shown (p < 0.0167). $\eta^2_{partial}$ is a measure of effect size, or proportion of variance explained by the effect.

Effect	Best vs. Comparison		Best vs. Neutral		Neutral vs. Comparison	
	F(1,39)	$\eta^2_{partial}$	F(1,39)	$\eta^2_{partial}$	F(1,39)	$\eta^2_{partial}$
bright - dim	14.49	0.27	37.78	0.49		
uniform - non-uniform	6.98	0.15	40.69	0.51	104.50	0.73
interesting - monotonous	8.23	0.17	107.47	0.73	43.71	0.53
pleasant - unpleasant	43.51	0.53	62.22	0.62		
comfortable - uncomfortable	41.38	0.52	33.30	0.46		
stimulating - subdued			67.38	0.63	27.09	0.41
radiant - gloomy	10.77	0.22	74.89	0.66	10.97	0.22
tense - relaxing	33.05	0.46			8.47	0.18
dramatic - diffuse			56.96	0.59	66.91	0.63
spacious - cramped	25.25	0.39	11.32	0.23		
glaring - not-glaring	15.18	0.28	12.68	0.25	50.43	0.56
friendly - hostile	39.70	0.50	25.37	0.34		
simple - complex	9.22	0.19	13.26	0.25	40.05	0.56
formal - casual						
realistic - unrealistic	16.12	0.29				

Overall MANOVA (Best vs. Comparison): Wilks' Λ = 0.306; $\eta^2_{partial(ave)}$ = 0.28; F(15,25) = 3.79; *p* < 0.01
Overall MANOVA (Best vs. Neutral): Wilks' Λ = 0.155; $\eta^2_{partial(ave)}$ = 0.41; F(15,25) = 9.11; *p* < 0.01
Overall MANOVA (Neutral vs. Comparison): Wilks' Λ = 0.142; $\eta^2_{partial(ave)}$ = 0.27; F(15,25) = 10.07; *p* < 0.01

Table 3. Results of factor analysis on semantic differential appearance ratings for Best, Comparison and Neutral images, for participants doing the experiment for the first time (n=40). Also shown is a similar analysis from Veitch & Newsham[15]. The number in the table indicates the factor on which the scale loads with an unrotated factor loading of >0.5. In cases where a scale loads on more than one factor, the factor with the highest loading is listed first.

ADJECTIVE PAIR ↓	Veitch & Newsham	Best	Comparison	Neutral
friendly-hostile	n/a	1	1	1
pleasant – unpleasant	1	1	1	1
interesting – monotonous	1	1, 4	3	1
comfortable – uncomfortable	1	1	1	1
subdued – stimulating	1	4	2	1
gloomy – radiant	1	4, 2	1, 2	1
spacious – cramped	1	1	1, 2	3, 1
tense – relaxing	1	4, 1	1	1
dramatic – diffuse	1	4	2	2
nonuniform – uniform	2	2	2	3
complex – simple	2	2	3, 1	2
bright – dim	3	3	1, 2	1
glaring – not glaring		4	2	2, 4
formal – casual	4, 3		3	3, 4
realistic – unrealistic	n/a	1	4	1
% Total Variance explained	63	73	74	73

Table 4. Luminances derived from the 40 Best images, for participants doing the experiment for the first time. Values for the six directly manipulated surfaces are shown, as well as combinations of surfaces, and important ratios. For combinations of surfaces the mean is weighted according to the number of pixels in the image representing each surface.

			Median	Min.	Max.
Left Partition	(A)	cd/m^2	52.1	16.5	106.3
Right Partition (not under bin)	(B)	cd/m^2	58.4	14.5	110.3
Partition (under bin)	(C)	cd/m^2	45.9	10.1	85.7
Desktop	(D)	cd/m^2	46.5	13.0	110.3
Distant Wall	(E)	cd/m^2	62.5	11.1	108.3
Ceiling	(F)	cd/m^2	80.5	17.2	113.3
mean of ABCD		cd/m^2	47.8	29.4	91.7
mean of ABC		cd/m^2	48.0	15.4	100.8
mean of AB		cd/m^2	50.4	16.4	107.5
A:B			0.96	0.25	6.03
B:C			1.24	0.19	3.40
D:ABC			0.77	0.20	5.63
AB:E			1.00	0.16	4.26
E:F			0.85	0.12	4.69
VDT screen:ABC			1.04	0.53	3.51
VDT screen:D			1.08	0.45	3.85
AB:F			0.74	0.18	2.00

Table 5. Results of the HLM analyses. All intercepts and slopes shown are statistically significant (p < 0.02)

Outcome	Predictor	Intercept	t (d.f. = 39)	Slope	t (d.f. = 39)	Proportion of Variance Explained
F_NONUNIFRM	RMSLNUNI	8	*2.54*	0.64	*9.14*	0.52
F_DIM	L*	109	*9.46*	-1.16	*-5.55*	0.27
	L_F	76	*11.89*	-0.48	*-4.94*	0.22
	RMSLNUNI	63	*12.43*	-0.35	*-3.47*	0.13
F_UNATTRAC	L_F	63	*13.28*	-0.29	*-4.03*	0.18

Note. In this analysis the proportion of variance explained refers to variance at level-1, at the level of the individual ratings. The total variance at level-1 is calculated using a 'random intercept model' that is, an HLM model with no predictors; call this σ_1^2. We then add the level-1 photometric predictor, which reduces the unexplained level-1 variance to σ_2^2. The proportion of variance explained at level-1 is then $(\sigma_1^2 - \sigma_2^2)/\sigma_1^2$.

Figure Captions

Figure 1. The base image used in the experiment, and the six surfaces that were independently manipulated in luminance.

Figure 2. Two example combinations of surface luminances. Below each is the binary phenotype that represents the combinations of luminances, made up of 5-digit genes for each surface.

Figure 3. Crossover and mutation from parents' phenotypes (see Figure 2) to create a son phenotype; mother's genes are underlined. Resulting combination of surface luminances is shown.

Figure 4. Overall flow diagram of software used in experiment

Figure 5. Interface for the experimental task. Participants rated the image for overall attractiveness. They then used the boxes at the bottom to indicate their brightness preference for each surface.

Figure 6. Experimental set-up. Participants viewed the projected image through a viewport (photo on left). The space into which they looked was black except for the projected image (photo on right, with front wall removed). Diagram shows side elevation, approximately to scale.

Figure 7. Calibration of projector screen luminance vs. image pixel grey level. Open squares show data from Feb. 2002, closed squares show data from June 2002.

Figure 8. Histogram of the attractiveness rating of the Best image for participants doing the experiment for the first time (n=40).

Figure 9. Histogram of the preference for change in brightness of the six surfaces in the image, for the Best image, for participants doing the experiment for the first time. Data for all surfaces are grouped (n=235).

Figure 10. Example from one participant of a Best image on the left (in this case, given an attractiveness rating of 10) and its Comparison image on the right (which received an attractiveness rating of 4).

Figure 11. Comparison of mean semantic differential appearance ratings for the Best, Comparison and Neutral images, for participants doing the experiment for the first time (n=40). A line between two symbols indicates a significant univariate difference (p < 0.0167) between those two image types for that adjective pair.

Figure 12. (Approx.) Average preferred luminance.

Figure 13. Preferred brightness in the field of view, as a percentage of the maximum available, for this study and a study done in a similar real space (Veitch & Newsham[16]).

Figure 14. "Spaghetti" plot of 40 individuals' subjective ratings of uniformity (F_NONUNIFRM) vs. a photometric descriptor of nonuniformity (RMSLNUNI). The best-fit linear regression line is superimposed.