Whisper, Don't Scream: Grids and Transparency

Lyn Bartram and Maureen C. Stone

Abstract—Visual elements such as grids, labels, and contour lines act as reference structures that support the primary information being presented. Such structures need to be usefully visible, but not so obtrusive that they clutter the presentation. Visual designers know how to carefully manage transparency and layering in an image to balance these elements. We want the presentation of these structures in complex, dynamic, computer-generated visualizations to reflect the same subtlety and comfort of good design. Our goal is to determine the physical, perceptual, and cognitive characteristics of such structures in a way that enables automatic presentation. Our approach to this problem does not try to characterize "ideal" or "best," but instead seeks boundary conditions that define a range of visible yet subtle legibility. All presentations that are clearly bad lie outside of this range, and can easily be avoided. In this paper, we report three experiments investigating the effects of grid color and spacing on these boundary conditions, defined by manipulating the transparency (alpha) of thin rectangular grids over scatter plots. Our results show that while there is some variation due to user preference and image properties, bounding alpha allows us to reliably predict a range of usable yet unobtrusive grids over a wide variety of conditions.

Index Terms—Information visualization, automated presentation, applied perception, visual design.

1 Introduction

Visual elements such as grids, labels, and contour lines act as reference structures or visual metadata that augment the primary information being presented. Such structures are meant to support the information presented rather than be part of it. They need to be usefully visible, but not so obtrusive that they clutter the presentation.

Visual designers carefully manipulate the visual balance between these different elements in the image, creating an attention hierarchy that reflects the information content of the presentation. Creating this balance is often difficult and time-consuming, even for a static image. In dynamic, computer-based visualizations, where the amount and type of information in the image is constantly changing, it is not practical to hand-craft each new presentation. The overall goal of our research is to understand and quantify these subtle aspects of visual representation such that they can be algorithmically manipulated to match human requirements in interactive and dynamic conditions.

Our broad interest is improving the appearance and usability of rich, complex, computer-generated visualization. The trend in interactive visualization has been to provide the user with more and more tools to manipulate the appearance of the image and to manage all the visual elements. Instead, we believe that we can exploit the capability of the visual system to extract visual information as needed and ignore it when irrelevant. By adjusting focus and attention, the viewer can see what is needed, and avoid distraction from what is not. Aside from the inherent elegance of this approach, it is substantially more efficient than dynamic manipulation [40]. What we seek are the physical and psychophysical rules to render information legible on demand; simply put—do it with your eyes and not your hands.

Our approach to this problem is not to characterize "ideal" or "best," but instead to define boundary conditions outside of which the presentation is clearly bad. We reason that the best solution will always be contextual, as well as a matter of taste. Boundary conditions, however, are more likely to have simple rules that can easily be incorporated by engineers and researchers, and less likely to be influenced by taste.

The experiments reported in this paper represent our first results toward characterizing the properties of subtle visual representation. Specifically, we look at overlaid grids, one of the most common reference structures. We present data to support the existence of an effective display range, described in terms of transparency (alpha), for thin rectangular grids over scatter plot data. Our experiments were structured to independently determine the two boundaries of the range. The lower boundary was defined to be the faintest usable grid; the upper boundary defined the point where the grid becomes too strong, or intrusive.

Our results show there is a useful range between the two boundaries that varies with image density, with a common overlapping region across all of our experimental conditions. As a result, we can recommend setting alpha for the grid to lie between 0.1 and 0.45, higher for dense plots, lower for sparse ones. We also demonstrate that for these specific examples, a value of 0.2 would be satisfactory for all conditions.

In this paper we present the motivation for the experiments, the experimental design for each of our three experiments, an analysis of their individual results, and

- *L. Bartram is with the School of Interactive Art and Technology, Simon Fraser University, 250 10345 012nd Ave., Surrey, BC, Canada V3T 0A3. E-mail: lyn@sfu.ca.*
- *M.C. Stone is with StoneSoup Consulting, 15340 162nd Ave NE, Woodinville, WA 98072, and the School of Interactive Art and Technology, Simon Fraser University, Canada. E-mail: stone@stonesc.com.*

Manuscript received 17 Jan. 2010; revised 23 June 2010; accepted 9 Aug. 2010; published online 29 Oct. 2010.
Recommended for acceptance by R. Balakrishnan.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCG-2010-01-0009.
Digital Object Identifier no. 10.1109/TVCG.2010.237.

how we how applied this analysis to create our recommended range.

2 Exploring the Visual Middle Ground

2.1 Perception and Attention

We can think of this quality of "legible only when needed" as a property of visual attention: why does a well-designed grid seem to be more visible when we pay attention to it, and what exactly does that mean? The art psychologist Gombrich describes a visual *middle ground* [16] where features can be "extruded" into the foreground or "receded" into the background by slightly changing the degree of attention. Previous theories of visual attention as a "spotlight" could not explain this, but recent research in task-directed vision and attentional effects on visual acuity promise some perceptual and cognitive ground for these effects [13], [15]. The theory of active vision suggests that a top-down, task directed process directs attention to the grid such that it moves up or down in an "attention scale." When it is not attended to, it becomes less salient. Once the object has been attended to, it may be subsequently more visible. Gobell and Carrasco reported attentional effects on visual acuity that include increased sensitivity to contrast and higher spatial resolution [15].

This suggests that elements in the image such as grids can be subtly set at levels that support this kind of "information on demand." We are interested in finding out indeed how subtly we can present such elements. More specifically, we seek a way to characterize these design decisions in terms of quantities that can easily be computed from computer displays. That is, we seek robustly validated metrics and principles for use for "legible, but not obtrusive."

We informally characterize this as a new metric, called the *JAD*, or *Just Attendable Difference*. Our working definition of *attendable* is as follows: phenomena that are *attendable* but currently not *attended to* exist as visual artifacts that are detectably part of a scene but can be uninvolved (i.e., not always visually salient) in the effort of interpreting that scene. It is this notion of attendability, we believe, that may carry with it the dimension of subtlety and richness that is key to the efficacy and utility of design and not yet fully integrated into the field of computer-generated visualization.

A JAD is similar to the just noticeable difference (JND) used in perception in that it is a uniform metric for visual differences. However, instead of being at the threshold of perception, it is a larger, more robust unit that quantifies subtle yet significant differences useful for layering and legibility. A JND is the difference between two stimuli that (under properly controlled experimental conditions) is detected as often as it is undetected. In contrast, we think of a JAD as the difference between two elements, or between an element and its background, that is only noticed or remarkable when visually needed (i.e., when attention is directed to it) but that is relegated to the background otherwise. We believe that attendable (JAD) is bigger than simply perceptible (JND), but measuring it is more challenging: attention is more complex than perception, and it introduces questions of aesthetics and utility, core to the emerging area of computational aesthetics in visualization.

2.2 Why Grids?

One of the most ubiquitous reference structures is the grid. It is essential in most two-dimensional representations where relative location is important, particularly maps and plots of various kinds. A grid has both local and global presence: the lines need to be appropriately visible at the region of interest, but the global structure of the grid can enhance that visible quality because the eye can predict where it "should" be (the Gestalt principle of continuity [40]). However, lines drawn across a busy image can interfere with the actual data representation and make the image itself very cluttered. While designers and cartographers carefully craft representations with well-balanced grids, many computer-generated representations have grids that seem either too intrusive or too faint. Excel in Office 2007, for example, creates default grid lines that are solid black on a white background. An experienced designer would make them light gray in proportion to the amount and type of information displayed.

3 Related Work

3.1 Design Principles

Designers create subtle reference structures by varying visual parameters such as color, contrast, and transparency to manipulate the Gestalt principles of figure and ground [40]. The overall goal of the designer is to achieve a well-balanced composition of visual layers, in which whatever constitutes the "figure" is well defined with respect to "ground." Grids and other visual metadata live somewhere in the middle of these layers, where sometimes the grid needs to be more *figure* (visually accessible for search or reference) and sometimes more *ground* (relegated to the background and not intrusive).

Designers approach the problem of visual complexity by carefully constructing an image from well-balanced layers [38]. They work to balance the visuals through well-understood design principles of hierarchy of information with formal principles, such as line weight, contrast, color, and texture. Factors such as the possible contexts of use are taken into consideration, including the thresholds of display technologies and methods of reproduction.

Designers add gridlines to help users to interpret and interpolate data. This is especially true for graphs that serve as lookup tables. At the same time, grids can be the source of distraction for seeing data. Designers use both generic and custom solutions to strike a balance between the usefulness and the distraction from grids. One example for generic solutions is the use of two different line weights for major and minor gridlines. Another generic solution is the use of gray gridlines with dark data plots. Custom solutions include adjusting the darkness of the half-toned grid, according to the type of plots, the amount of data, and background color.

With respect to data visualization, a common rule of thumb is to minimize the amount of "ink" used on nondata information in visual displays of data [12], [38]. Tufte shows examples of removing unnecessary elements from data displays, even to the point of creating the illusion of a white grid by erasing lines through bar charts [38].

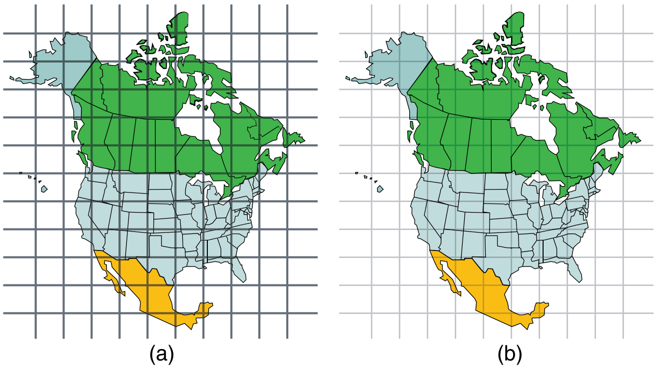


Fig. 1. (a) A badly designed grid that obscures the underlying information. (b) This grid is more subtle, allowing the viewer to focus on the map.

Fig. 1 shows a grid overlaid on a map. The grid lines in (a) are bold and dark, creating a highly visible structure in the foreground. In (b), the grid lines are thinner, lighter, and rendered transparently using alpha blending. As a result, the grid blends with the map image in a way that makes it visible, but unobtrusive. The use of transparency is critical for this integration, as it makes the lines adaptively darker than the background information. Making the grid a constant light gray would create a structure that alternated between being lighter and darker than the underlying graphic and make the grid itself less coherent. This use of transparency for grid design was the inspiration for our experiments.

3.2 Vision Research in Transparency

The perception of transparency that enables the visual system to separate a scene into overlapping layers is clearly important to this work. Simply put, perceptual transparency is a type of surface perception that illustrates the visual system's remarkable ability to reconstruct the three spatial dimensions of the environment given an image with only two [8]. Brill suggests that the appearance of transparency is critical to creating an effective grid (or indeed most reference structures or distinct objects), because it allows *color scissioning* [7]. This means that the visual system perceives the stimulus as two layers, which can be separated and independently analyzed. He proposes that the degree of transparency is the critical perceptual factor in layering such elements.

How perceptual transparency actually works is a subject of active research in the vision community, from which two complementary theories have emerged: Mettelli's "spinning" disk, which uses a simple equation similar to the computation of alpha to define the relative reflectance of an area partially obscured by a transparent surface [29] and edge-based X-Junctions [8], which focuses on the relative lightness of regions around an X-shaped intersection of regions.

Vision research on transparency has concentrated on determining the principles that make areas of overlaid color appear transparent. Reference structures such as grids, contour lines, or user interface control points are constructed of thin lines that may be rendered using alpha blending, a technique already known to create the perception of transparency. For our purposes, what needs to be explored are the conditions in which transparency can be reliably manipulated to achieve both legibility and subtlety, especially in the conditions of thin lines and sparse structures.

3.3 Transparency Used in Visualization

Transparency has been applied somewhat sporadically in visualization as a representation dimension, notably to show uncertainty by making uncertain objects less opaque [10], overlaying a transparent wash for highlighting [31] and more generally for reducing screen space limitations by overlaying objects or features [27]. MacEachren and Kraak include it in a list of cartographic visual variables [27]. These effective levels of transparency are often not critically or systematically validated. Pertinent to our interests, a number of user interface techniques use transparent reference structures, notably for two-handed input tools [6], [14], [41]. Approaches by Bier et al. [6] and Fitmaurice et al. [14] implement overlay tools such as lenses and selection regions as semitransparent regions. A study by Zhai et al. [41] showed that rendering a 3D cursor as a semitransparent surface-aided selection and navigation more than its wireframe equivalent and did not distract from performance. In an effort to increase access to occluded windows, Ishak and Feiner [21] rendered the "unimportant" part of the overlapping windows semitransparent. While they did not empirically establish the optimal transparency level, they claim that a setting of 85 percent translucency was effective in maintaining the context of the overlapping window and the legibility of the overlaid window contents.

Harrison [17], [18] investigated legibility, attentional demand, and object identification in icon palettes and menus varying both the transparency of the menu surface and the complexity of the backgrounds on which they were superimposed (text remained opaque). In the text menu study, three types of background were used: text pages, wireframe images, and solid images, aligned such that a major part of the content was underneath the 12-item menu. Six transparency levels and two types of overlaying the font were tested. Text was opaque: only the surrounding label surface was altered. As might be expected, both transparency and image background were strongly significant. There were interactions between background and transparency (transparent menus over more complex backgrounds took longer to resolve). Overall, however, strong interference effects did not occur until transparency was at or greater than 75 percent, suggesting that even slight transparency is effective at partitioning objects from background. This was borne out by subjective reports from participants that even a tiny change in the transparency level (from 100 to 90 percent) made a huge difference in the ease of finding the right menu item.

3.4 Contrast and Legibility

A usable grid must be sufficiently visible to be legible. The visual system detects changes in luminance (perceived lightness) to define shapes and edges [40], [25]. Luminance contrast, computed from the relative luminance between a symbol and its background, has long been used to predict text and symbol legibility [23], [28], [4].

More recently, the quantity L^* has been used for evaluating text legibility [42], and has been found to give results consistent with luminance contrast. L^* is a perceptual metric for lightness (computed from luminance and a reference white) where numeric differences uniformly describe perceptual differences, and one unit is a minimally distinguishable difference [36], [40]. ΔL^* is an attractive metric for design because its perceptual uniformity makes it easy to interpret for all colors, and because of its integration with design software such as Adobe Photoshop.

Luminance, luminance contrast, and L^* are readily computable for colors defined on digital displays [36]. Therefore, evaluating grids and other reference structures in terms of luminance contrast or differences in L^* seems a useful approach, at least for determining the lightest usable values for the grid. Somewhat to our surprise, we found that our results cannot be simply described in terms of luminance contrast or L^* differences. These results have been reported in detail elsewhere [35].

3.5 Design Evaluation

As several researchers have pointed out, assessing the effectiveness of a visualization method is challenging. Acevedo summarizes that evaluation of visualization methods is typically either anecdotal, via feedback from or observation of scientific users, or empirical, via measurement of the performance of relatively naïve users on simple abstract tasks [1], [2]. In particular, the controlled studies that examine the efficacy of a particular technique have a number of drawbacks that limit their utility [1], [37], as they require substantial effort to execute, involve nonexpert users and constrain the scale of conditions to assess. Thus, their results are sometimes difficult to generalize to different, more complex environments and tasks.

Recent studies by Acevedo et al. [1], [2] and Tory and Möller [37] employed the inclusion of critiques from experienced visual designers (an established method in design) both as a rich and effective evaluation method and as a way of increasing knowledge to guide the creation of new visualization methods. In [1], Acevedo et al. combined perceptual principles (choosing a set of visual dimensions corresponding to perceptual features) with design reviews, where a set of experienced visual designers critiqued a wide set of simple visualizations and suggested redesigns. In [2], they used the same approach and correlated the design ratings with the results of a concurrent quantitative experiment. They verified that the design ratings largely cohered with the more limited quantitative results. They conclude that expert design reviews are a critically underutilized method of evaluating both the effectiveness and the appropriate use of visualization techniques and advocate that in many cases this approach can take the place of more limited, empirically-based methods with visually less literate users.

4 Approach

Vision research gives us some ideas about the underlying metrics we might examine for how a transparent grid should be useably detectable: in other words, how faint it can be. It provides little insight, however, into what makes a grid move from comfortably usable to too intrusive. Design practice, on the other hand, tells us how a grid should be subtly layered against the background, but does not explain the underlying operating principles for why it works in any but the most general terms. More specifically, experienced visual designers can critique existing approaches and advocate how to improve them (e.g., "A grid should always be well balanced with respect to both the foreground figure and the background"). However, to adapt to dynamic reconfigurations of that visualization the underlying visual elements that influence that balance, such as complexity, contrast, or legibility, must be algorithmically identified and manipulable. Designers cannot tell us from that perspective what these critical factors and principles are.

It is those principles, or at least a robust approximation of an implementable model, that we are seeking. We therefore took an empirical approach that was informed by ongoing consultations with expert designers. The objectives were twofold. First, we wanted to discover if there was general agreement of the best levels of a grid and what they were. Second, we were curious about how preferred grid settings would change for different types of images and backgrounds, as we conjectured that contrast with the background would be a strong determining factor.

We began with careful consultation with our designer colleague, Dr. Diane Gromala, around the practice-proven principles of careful grid design and visual layering. There were two key results of this consultation. The first is that well-designed grids were far more visually subtle than common in visualization systems. That is, on a white background, a designer would create a grid that appeared a very pale gray, but many application developers program grids that are black. The second is that digital designers achieve this subtlety using transparency (alpha blending).

Based on these insights, we created a series of pilots, which we performed at several conference poster sessions. We recorded not only the numeric results but also collected comments from our users to help us refine our experiments. The conference setting also allowed us to discuss our goals with a variety of visualization researchers and designers. This led us to our experimental design, which asks users to rate the quality of a grid displayed in a familiar context (scatter plots) where a typical use (determining scale values) is implied.

As Acevedo et al. have pointed out [3], the trade-off in experiments versus design preferences has to do with both ecological validity (how much the task or condition relates to "real-world" situations), scope (how well a solution generalizes to other conditions) and rigour (how thoroughly can you substantiate your assertions about standard knowledge or practice?). How to determine the "minimum" level of a grid is straightforward with experimental methods from applied perception. However, methods to determine the best setting for a grid—where "best" may be something more than minimal and less than full force—introduce questions of aesthetic judgment and perhaps individual preference. An intrusive grid can impede legibility and utility of the visualization it is meant to support, but designing the tasks to elicit these performance metrics is not straightforward. We could use very simplistic tasks related to visual search and legibility, but these would not elicit the subtler aspects of ease of use over time and user comfort that we believe contribute to a good grid. Such questions are related to the *quality* of a grid. As Norman points out, while quality cannot be explicitly identified in usability, its importance in design is uncontested and people inherently like using well-designed things better [30].

We believed that visualization users would reasonably judge grid settings according to how they would best serve a visualization, as grids are so familiar, so we elected to observe subjective judgments of optimal grids rather than task performance as our experimental measure.

We used images chosen from typical grid applications (scatter plots and maps). Using these familiar applications introduced a familiar task context without task performance. Finally, we chose participants with a variety of visualization and visual expertize as our experiment subjects. Our reasoning was that while visual design experts could determine what a grid "should" be according to their training and practice, people experienced with using grids in dynamic visualization contexts might vary in their preferences. Because we chose familiar visualizations for our grid applications, we could also use "naïve" users, who would be familiar with using plots in Excel, for example, or road maps.

To summarize, we used designers for guiding the initial choices of what to set for grids, and participants across a variety of visual and visualization experience to manipulate these settings according to different images.

4.1 Pilots

To begin exploring this problem, we created an interactive tool in Adobe Flash that allowed us to change the grid parameters: line width, gray value. and transparency. Our initial image was the same map as in Fig. 1 rendered in shades of gray. Based on our own interactions with this tool, plus data collected informally from a handful of colleagues at SFU/SIAT, we developed a Flash tool for our pilot studies. The first of these were held on two contiguous days at the ACM Applied Perception and SIGGRAPH conferences, in the poster sessions. The second set was conducted at the IEEE Visualization conference in the interactive demo session. The UI for the tool and the instructions were changed slightly between the two sessions, in response to user comments. The users were instructed to manipulate the color (gray pixel value) of the grid, and its transparency (alpha value). Each participant was asked to produce three grid settings for each image according to the following descriptions:

- **Best**: Please adjust this grid to what you consider is optimally usable.
- **Faintest**: Please adjust this grid to be as faint as you think it can comfortably be to be still useful; any fainter and you would no longer be able to easily use it.
- **Strongest:** Please adjust this grid to be as strongly visible as you think it can comfortably be before it interferes with or "comes in front of" the image; any stronger and it would be too obtrusive.

We collected as data both the gray scale and alpha settings as well as comments from the subjects. We used two different laptops, with significantly different "gamma" settings (1.8 and 2.2) to explore the effect of the display on the results. Both displays were calibrated prior to use, and allowed to warm up for at least hour to stabilize.

Our key insight from the pilots was that the concept of "best" was highly variable, but that the range within which this value of best could be found seemed surprisingly consistent. This led us to focus our formal experiments on determining two boundary conditions for a transparent grid of a fixed line weight and spacing: the point at which it is imperceptibly light (too *faint*), and the point at which it clearly sits in front of the image, rather than seeming a part of it (too *strong*). It appeared than an ideal grid sits between these boundaries.

Our second insight was that the visual effect of the strong grid was qualitatively different from the faint. Several of our participants called the strong grid a "fence." We were struck by their descriptions of how there was a particular point where the grid would appear to "detach" from the image and pull away from it or in front of it. This led us to include the "comes in front of" language to the second set of pilots and in our formal experiments.

We found that the display gamma did influence the choice of alpha values, in that separating the data by display created a cleaner and more consistent set of boundaries. Therefore, we decided to control this variable in our formal experiments.

User feedback was generally positive. Users found the task interesting and not too difficult. We received constructive feedback on how to simplify and improve the user interface, which we incorporated in our formal studies. Most critically, users found that adjusting both gray and alpha unnecessarily complex, so our formal studied used only alpha as a variable.

5 The Grid Studies

We designed a series of three experiments to see how accurately we could predict the faint and strong alpha boundaries. Participants were asked to adjust the alpha value of a grid with a constant line weight of one pixel and a constant color (black or white) over a set of images with different background colors (gray values), and different levels of visual complexity (plot density and grid spacing).

We collected two types of data measures: the alpha settings for each boundary, and the range between them (i.e., for each subject and condition, we calculated the difference between the mean alphas for those boundaries). We investigated the effects of grid color, image background, plot density, and grid spacing on the two boundaries and the range. We used the same method and metrics in all three experiments.

5.1 Method

We set the participants to two different tasks. The first was to specify the point where "the grid is useably perceptible without being unnoticeable"(*faint grid*). The second was to adjust the grid "to meet your best judgment of how obvious it can be before it becomes too intrusive and sits in front of the image; some users have called this a fence" (*strong grid*). It is important to note that this terminology came from observations of previous participants in pilot studies describing each of these effects; we simply repeated it. We explained that we

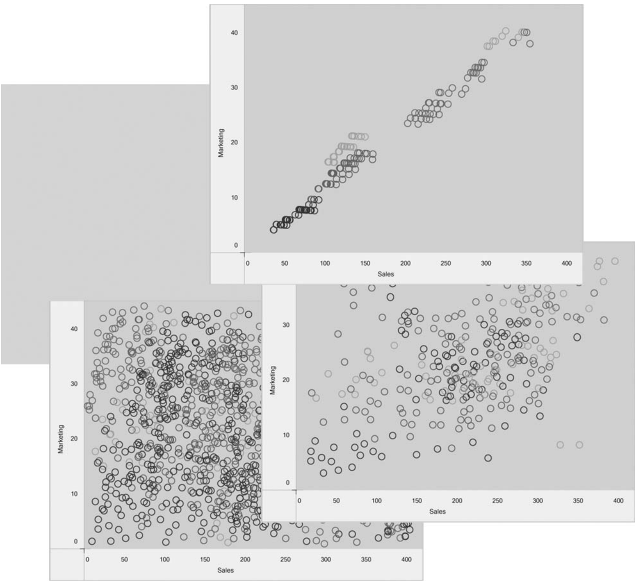


Fig. 2. The images used in the experiment, showing the four levels of density on one of the gray backgrounds used in the experiment.

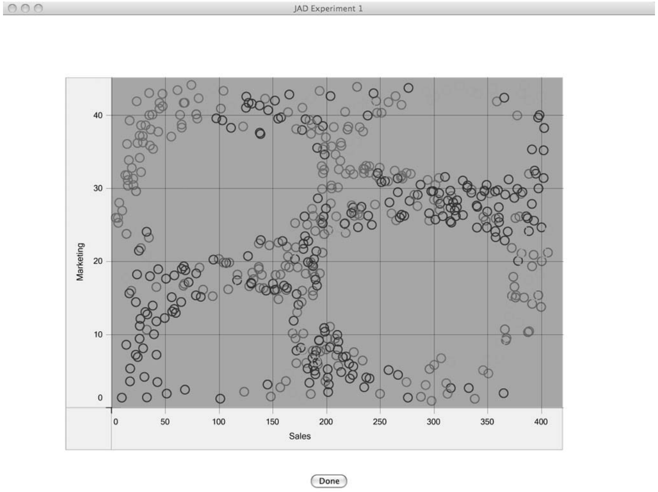


Fig. 3. The experimental setup, showing one of the training images. The user manipulates the alpha values with the mouse buttons, then presses the small button "Done."

were looking for grid settings that were "still useable": that is, each may have been at the border of tolerable, but still on the acceptable side of the boundary. Finally, we also emphasized that there was no right answer and that we were collecting user preferences rather than performance data.

Using a standard computer monitor (an Apple Cinema display), the participant was presented with a series of images. For each image, (s)he would adjust the grid transparency to satisfy the task (faint or strong grid). By providing only one variable, we could create a relatively simple interaction based on the motion of the mouse. Holding down the left mouse button increased the strength of the grid (increased alpha); holding down the right button made the grid fainter (decreased alpha). Therefore, the user could make the grid fully transparent by holding down the right button until the grid alpha became 0; alternately, he or she could turn the grid full "on" by holding down the left button until the value of alpha became 1. Once either of these limits was reached there was no grid change. There was no time constraint on the task, and participants could play around with the settings as much as they liked until they were comfortable with the result. The mouse was active any time it was over any part of the image although location in the image had no effect.

The participants performed the tasks as two separate tests; that is, they did all of one task on all of the images, then the other task. Participants could practice on a set of training images for an unlimited time although, in practice, all users were comfortable with both tasks after a few practice images. All users performed the experiments on the same, calibrated display under the same viewing conditions.

Underlying all these studies is the fundamental hypothesis that alpha provides both a reasonable tool for adjusting grid appearance and a potentially interesting basis for modeling adaptive representation.

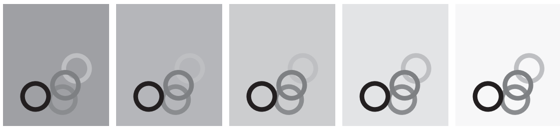
5.2 The Images

While our pilots used both maps and scatter plot images, we used only scatter plots in our formal experiments because they were easier to construct and manipulate to create variations in image complexity. They are familiar, and the use of grids is essential in understanding them. Therefore, although we were not asking our participants to carry out performance-based tests on actual grid comprehension, we introduced at least an ecological context of use.

Our pilot experiments suggested that image complexity had a noticeable effect, especially on the strong boundary of the grid. Image complexity is a broad term [34], encompassing different definitions and levels of detail such as pixel coverage, spatial frequency, color palettes, segmentation, number of visual elements or number of different visual codes (the latter two can be considered measures of information complexity). We are interested in exploring the implications of these various dimensions in further studies with respect to the overlaid reference structures, but for the purposes of these studies, we began with a simplistic approach to complexity, in which we varied the amount of background covered by a small set of visual elements. Our hypothesis was that for images that were predominantly background, we would see a fairly simple relationship between luminance contrast with the background and the choice for the grid boundaries, similar to the legibility criteria for text and symbols.

We created four image types of varying density: a flat field (which can be considered "no density") and three scatter plots at different levels of background coverage: sparse, medium, and dense (Fig. 2). The plots were generated by creating a dense scatter plot, then hand manipulating it to affect the distribution and the number of elements. The gray values of the foreground circles in the plots were chosen to supply some visual variety across a range of lightness levels, and to be visibly different from all of the different background levels.

Each plot was rendered as a JPG image and displayed at a spatial resolution of 800×600 pixels on an Apple Cinema Display (liquid crystal), as shown in Fig. 3. The starting value of alpha was set at 0 (i.e., no grid) for both the faint and the strong tasks. The user manipulated the grid until



L* value	60	69	78	87	96
PC pixel	143	167	191	217	243
Mac pixel	126	152	180	209	241

Fig. 4. Dark Grid background gray values, Experiments 1 and 3.

satisfied, then pressed the button in the bottom center of the screen labelled "Done."

5.3 Display Calibration and Viewing Environment

All experiments were carried out on the same display in the same dimly lit room. The display was calibrated using a Gretag Eye-One spectroradiometer (10 nm resolution), and Gretag's profiling package to achieve a gamma of 1.8 (typical for Macintosh systems) and the native display color temperature. Precise color specification was not important for this experiment, as images and grids were all in shades of gray. Subjects sat 59 cm (24″) from the display, which had a spatial resolution of $1{,}920 \times 1{,}200$ across a 23″ diagonal viewing area. All grids displayed were a single pixel wide, or 1.5 minutes of arc. The images were displayed with a large white border to control adaptation.

5.4 Hypotheses

We had several hypothesizes in these experiments, based on our experience with the pilot studies.

- H1. The faint boundary for the usable grid would show less variation than the "fence" setting. We conjectured that the faint grid setting is constrained by minimum perceptibility, while personal taste would play a much larger role in the judgment of intrusiveness.
- H2. Alpha for the faint setting would be less than 0.5. The common practice in design is to set transparent structures at this level or lower; we wanted to test this assumption on subjects without design training.
- H3. Background would have an effect on alpha settings in both the faint and strong cases. We expected that less transparent grids would result from darker backgrounds (in the black grid case) and from lighter backgrounds in the white grid case.
- H4. Plot density would affect alpha settings. Our inspection of numerous gridded images suggested that dense images made it harder to see subtle grids.
- H5. Grid density would affect alpha settings. We surmised that the finer precision implied by a dense grid would result in a less transparent setting.
- H6. Results would be symmetric for the light and dark grids. We expected this result to emerge as a simple effect of luminance contrast.

6 Experiment 1: Black Grid on a Light Background

In our first experiment, we used a black grid ($\mathrm{RGB} = 0, 0, 0$) with a fixed spacing of 86.5 pixels in x and 118 in y, to align with the x- and y-axis values in the scatter plots.

Each image was displayed over 5 gray backgrounds, ranging in uniform steps from $\mathrm{L}^* = 96$ to 60 (Fig. 4).

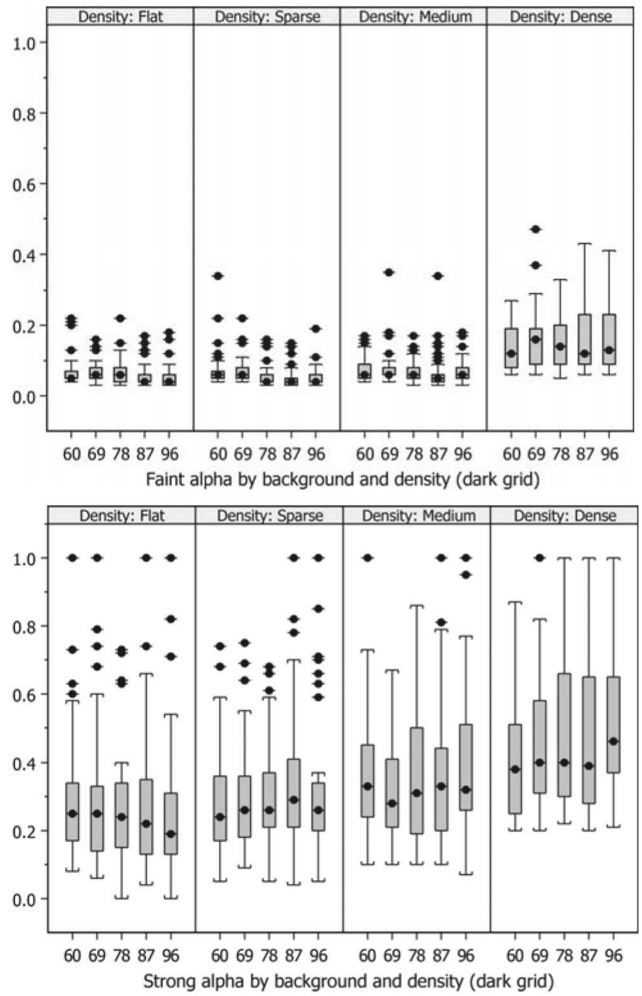


Fig. 5. Alpha by background and density for Experiment 1 (black grid).

6.1 Experimental Design

A $4\ (\text{density}) \times 5\ (\text{background})$ factorial design yielded 20 experimental conditions. We used a split-plot design in which each subject performed two separate task blocks, one for each grid condition (faint or strong). Each task block had three repetitions of 20 images resulting in 60 trials/block. Trial ordering was randomized and block ordering was counterbalanced. Thirteen university students with normal or corrected-to-normal vision participated in the experiment and were paid.

6.2 Results: Black Grid

The results can be seen in Figs. 5, 6, and 7. The key finding is that background is not significant, refuting **H3**. A simple ANOVA showed no significance of task block order in this or, in fact, in any of the subsequent experiments so we do not discuss this further.

A two-factor ANOVA revealed a significant effect of density in both grid conditions: $\mathrm{F}(3{,}236) = 60.0112$, $\mathrm{p} < .001$ (faint) and $\mathrm{F}(3{,}236) = 11.9789$, $\mathrm{p} < .001$ (strong), confirming **H4**. To examine this further, we iteratively removed different density data. When we performed the ANOVA on the flat and sparse plot data, we saw no significant effect of density in either the faint or the strong grid. Thus, only the medium and dense densities had an effect. Most

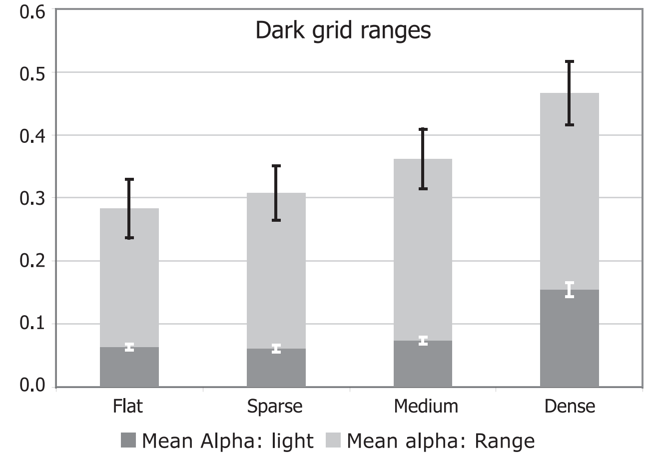


Fig. 6. Mean faint alpha and range as a function of density for Experiment 1 (black grid): the mean strong alpha is the sum of both.

subjects found the grid to be useably legible at very light alpha values, even for a dense plot. Even in the dense case, the faint boundary was less than 0.5 (confirming **H2**).

As we expected (**H1**), there was much less variation in the faint condition across subjects than the strong, where the data were noisier. However, even in the strong condition there was substantial agreement between most of the subjects, with only two outliers consistently setting a much higher level.

We also examined the individual *ranges* (the difference between faint and strong). The results are consistent with those for the grid boundaries. Background had no significant effect on range, but density was significant ($F(3,228) = 2.992,\ p < .03$). The range defined by our boundary conditions, which is plotted in Fig. 6 offset by the faint alpha, increases with density, as does the minimum alpha for the dense image. There is a rise between the relatively constant settings of the flat-medium densities and the dense case in terms of alpha value. However, as can be seen in Fig. 6, this effect is not large: the extent remains roughly similar. The previously mentioned two subjects who tended to set their strong boundaries higher caused significant subject variability in the range settings. When we removed these subjects' data from analysis, there was no significant effect of density. There were also no interactions between background and density.

In summary, Fig. 7 shows the distribution of mean alpha across all factors for this experiment: in other words, the simple means for each factor independent of the others. These graphs are useful for comparing the range of means across factors. As noted, the spread in mean settings is greatest across subjects, skewed by two outliers. However, this variability is only pronounced in the strong grid. Contrary to **H3**, background effect varied very little in both grid cases. Density means were similar with the notable exception of the dense plot (darkest line): this was the overall main effect.

7 Experiment 2: White Grid on a Dark Background

For Experiment 2, we replicated Experiment 1 except we used a white grid on five dark L^* backgrounds ranging from 4 to 60, as shown in Fig. 8.

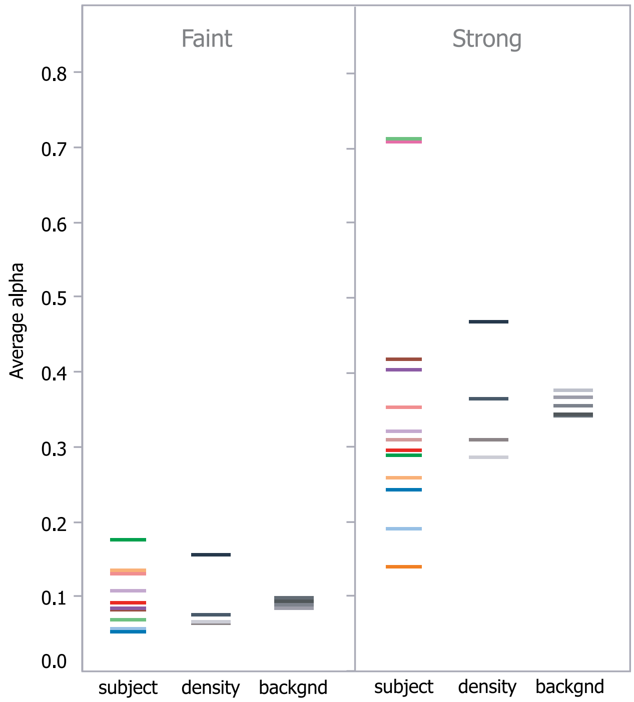


Fig. 7. Alpha distribution across factors, Experiment 1 (black grid).

The $L^* = 60$ case was included specifically to compare to Experiment 1, as this color will visually support either a light or dark grid. We gave subjects the exact same tasks and instructions.

7.1 Experimental Design

A $4\ (\text{density}) \times 5\ (\text{background})$ factorial design yielded 20 experimental conditions. Each grid task block had three repetitions of the 20 conditions resulting in 60 trials/block. Trial ordering was randomized and block ordering was counterbalanced. Fifteen university students with normal or corrected-to-normal vision participated in the experiment and were paid. None had participated in Experiment 1.

7.2 Results

The white grid results can be seen in Figs. 9, 10, 11, 12, and 13. As in Experiment 1, subjects set useably consistent values for both the faint and strong grid, with the setting for the strong grid more variable than for the faint one (**H1**). There was a significant effect of density (**H4**). However, refuting our hypothesis (**H5**), the results are not symmetric with the black grid, and the settings varied with background (**H3**).

For the faint case, both density $[F(3,332) = 87.82,\ p < 0.01]$ and background $[F4(4,232) = 14.3,\ p < 0.01]$ had

L* value	4	15	30	45	60
PC pixel	23	42	72	106	143
Mac pixel	14	28	55	87	126

Fig. 8. Dark background gray values, Experiment 2 (white grid).

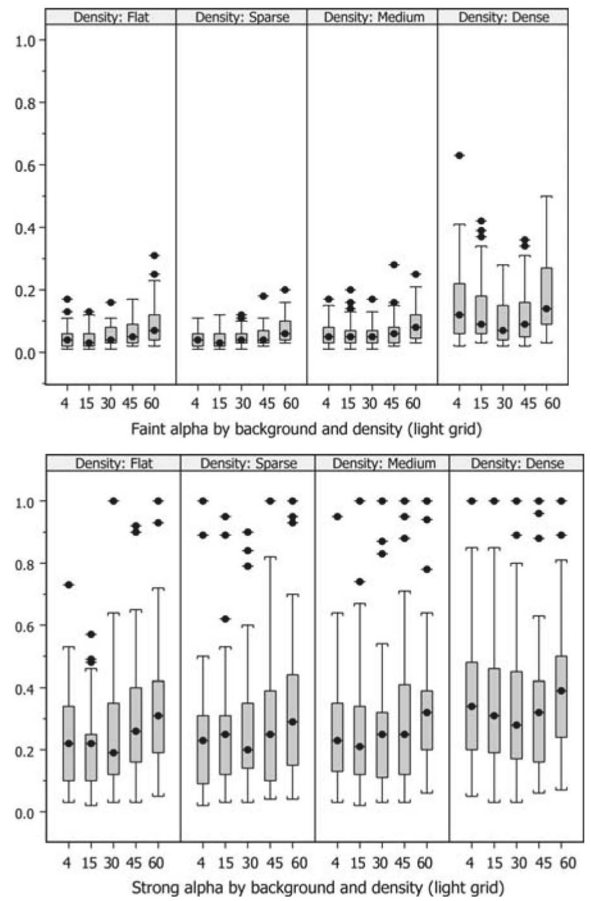


Fig. 9. Alpha by background and density, Experiment 2 (white grid).

significant effects in this experiment (Fig. 9). There was a significant interaction between them $[F(4{,}955) = 2.82, p < 0.01]$ (Fig. 10).

From Fig. 11, it can be seen that background had the strongest effect in the dense plot. Conversely, density was

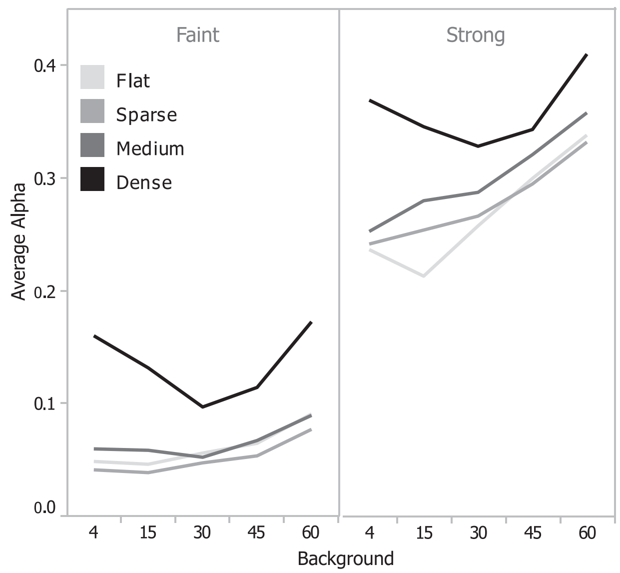


Fig. 10. Interaction between density and background for Experiment 2 (white grid). The different shape for the dense curve shows the interaction.

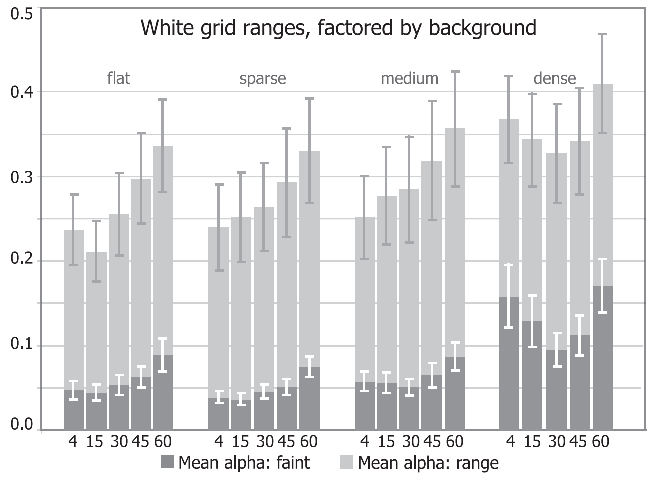


Fig. 11. Mean faint alpha and range as a function of background and density for Experiment 2 (white grid). There is a small but significant dependency on background for this case.

most significant in the darker backgrounds (4 and 15). The size of these effects was approximately 0.10 alpha. We note this is a perceptibly meaningful difference.

These effects together can also be seen in Fig. 10, showing that there is a small but significant dependency on background for the faint setting in this case. Again, there was notable subject variability in the strong settings, with most ranges between 0.3 and 0.5. However, a standard ANOVA showed were no significant effects on the range values themselves.

When we compared the mean alphas for the L*60 background, present in both experiments, we found that the grid (black or white) had a significant effect in the faint case (Fig. 12). While this difference was not large, it means that people set a different faint alpha for the same background for the black and white grids, and that they preferred the white grid more salient than the black. This effect was most apparent in the dense case. We saw no corresponding significant difference in the strong grid settings.

Fig. 13 shows the distribution across all factors for this experiment. As expected, we saw more variation in how subjects set the strong grid as opposed to the faint. In summary, compared to the dark grid settings of Experiment 1, we noted three differences. First, background had an effect.

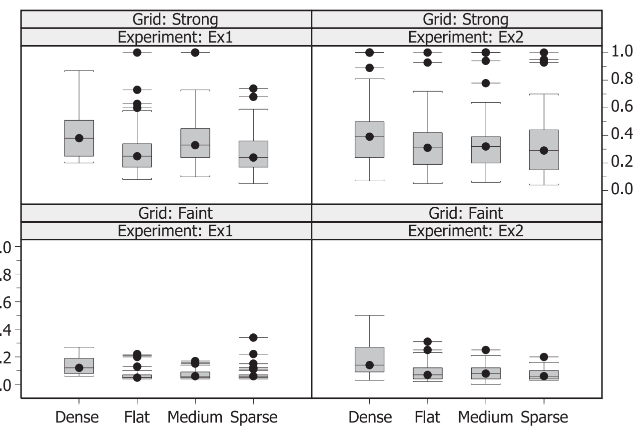


Fig. 12. Alpha settings for L*60 for black and white grids.

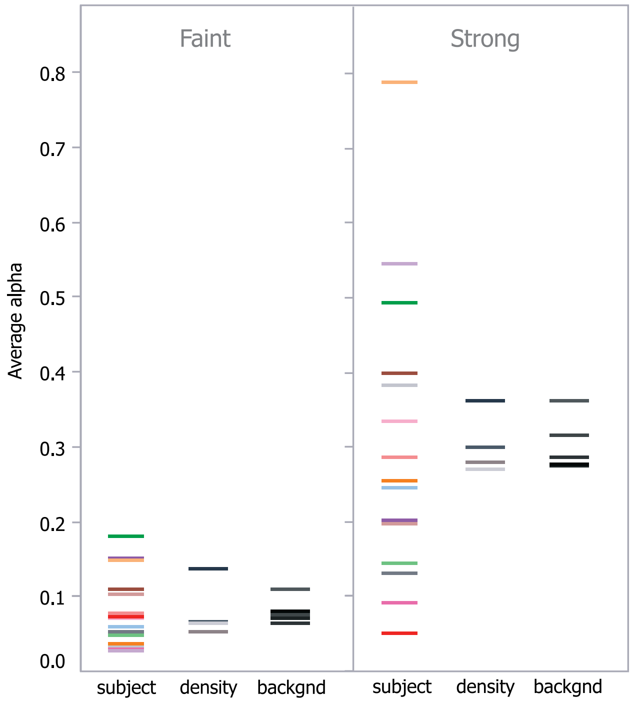


Fig. 13. Alpha distribution across factors, Experiment 2 (white grid).

Second, variability in strong settings was more pronounced among subjects for the white grid. Third, and most interesting, is the dissimilarity in both faint and strong means for the dense case. The mean for the strong setting is somewhat lower for the white grid than for the black. At the same time, the mean for the faint case with a similar background (L^*60) is higher for the white than the black. Clearly, our expectation of symmetry between the white and black grids (**H6**) is unfounded. We discuss this result further in Section 9.

8 Experiment 3: Grid Spacing

We then turned our attention to the effect of grid spacing, using a black grid on a light background, as in Experiment 1. Grid spacing varied from sparse to dense using 5 square spacings, expressed as the pixel difference in both x and y: 20, 40, 60, 80, and 100. These spacings are generally smaller than those in Experiment 1 (86.5 pixels in x and 118 in y). To reduce the number of conditions, and because background had not shown as significant in Experiment 1, we used only two of the background gray levels: $L^* = 78$ and $L^* = 96$.

Based on the results of Experiment 1, we expected to see little background effect. We hypothesized that the density of grid spacing, however, would have an effect analogous to plot density: that is, as grid density increased (spacing decreased), both faint and strong alpha levels would increase (**H5**).

8.1 Experimental Design

A 4 (plot density) $\times$ 5 (grid spacing) $\times$ 2 (background) factorial design yielded 40 experimental conditions. Each grid task block had two repetitions of the 40 conditions resulting in 80 trials/block. Trial ordering was randomized and block ordering was counterbalanced. Twelve university students with normal or corrected-to-normal vision participated in the experiment and were paid. None had participated in Experiment 1 or 2.

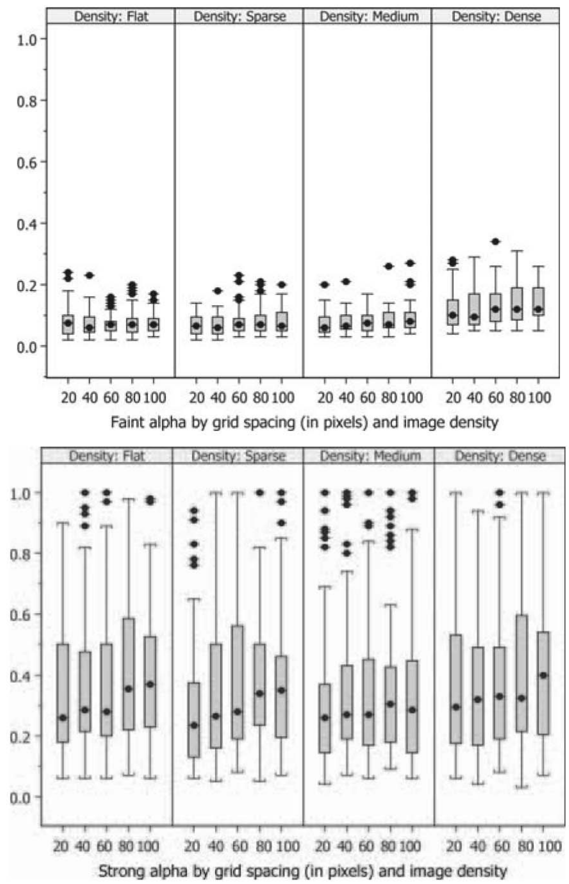


Fig. 14. Alpha by plot density and grid spacing, Experiment 3 (grid spacing). Spacing values are pixel separation.

8.2 Results

The results can be seen in Figs. 14, 15, and 16. Refuting **H5**, a primary analysis showed grid spacing had no significant effect in either the faint or strong case, although there was a mild trend of effect in the faint case when we considered it with respect to plot density: $F(4{,}955) = 2.32$, $p < 0.07$). There was, however, no significant interaction with plot density.

As in Experiment 1, plot density was overwhelmingly significant in the faint grid: $F(3{,}956) = 77.51$, $p < .01$. This was solely due to the effect of the dense plot: when we removed it from analysis, plot density became insignificant. As in Experiment 1, background was not significant. (Fig. 14)

The data were much noisier (**H1**) in the strong grid where subject variability was the overwhelming effect and we saw no single main effect for plot density, grid spacing, or background. Closer inspection of the data showed this was largely due to two subjects (S1 and S2) who consistently set the strong grid alpha quite high. When we removed their results from analysis, we noticed some interesting effects that partially supported our hypothesis about grid spacing. Plot density was again significant: $F(3, 796) = 4.7$, $p < .01$; but now grid spacing had an effect: $F(4, 795) = 3.68$, $p < .01$. There was no interaction between grid spacing and plot density.

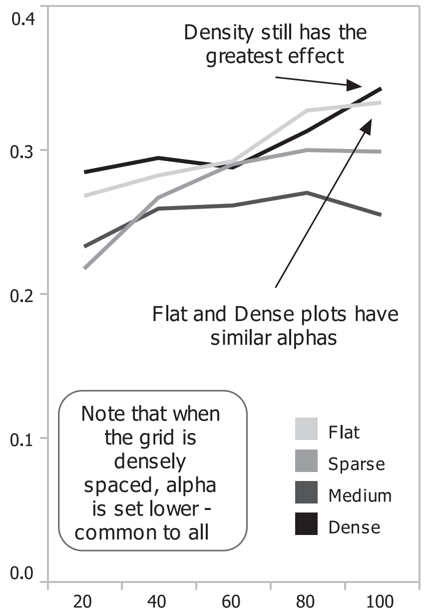


Fig. 15. Alpha as a function of spacing for the strong grid with outliers S1 and S2 removed.

Fig. 15 shows these effects in more detail. As we expected, alpha settings were strongest for the dense grid, but only marginally less than for the flat grid. These effects were most noticeable for the sparser grid densities (larger spacings). When the grid was more densely spaced the alpha settings were lowest (i.e., the grid was more transparent.) While this seemed to confirm **H5**—that spacing would affect alpha—it was a different effect than we anticipated. In fact, it appears from these results that while *plot* density increase encourages a stronger grid, increasing *grid* density has the opposite effect. We conjecture that this is due to the Gestalt principle of continuity—it is perceptually easier to "fill in the blanks" of a denser grid.

Fig. 16 shows the ranges for this experiment as a function of plot density and grid spacing (in pixels). The effect of plot density is similar to Experiment 1, and the overall boundaries are largely commensurate with the first experiment's results in the faint grid. We see the effect of grid spacing for the strong case, however, in the flat and sparse grids. This effect of grid spacing seems to have reduced the strong settings in this experiment, again (we conjecture) due to the continuity properties of the grid object.

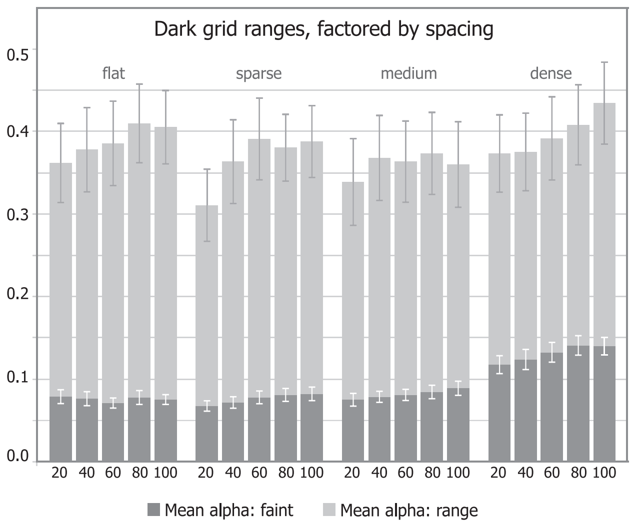


Fig. 16. Mean faint alpha and range, as a function of plot density and grid spacing, Experiment 3 (grid spacing). The mean strong alpha is the sum of both.

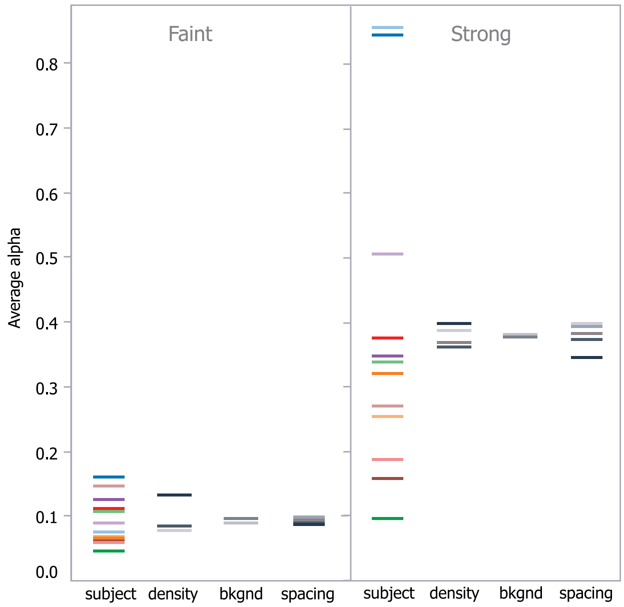


Fig. 17. Alpha distribution across factors, Experiment 3 (grid spacing).

Fig. 17 shows the distribution of responses across all factors. Plot density remains important. Grid spacing is significant in the strong case. We saw more variation in subject preferences for the strong grid in this experiment than in Experiment 1. There are trade-offs between grid and plot density settings, but these are dominated by plot density. We see a similar pattern to Experiment 1: while there are some subject outliers, the majority of the data sit closely around the overall means.

9 Discussion

Our primary result is that subjects set usefully consistent boundaries for these types of images. As expected, the results for the faint boundary are more consistent than for the strong boundary, and there was significant subject variation. Statistically, however, there were clear, consistent ranges between too faint and too strong, suggesting that establishing these types of boundaries may be a useful way to characterize subtle visualization.

Our subjects consistently set a faint boundary around $\text{alpha} = 0.1$, except for the dense plot, which is slightly larger. The results for the strong case show more subject variability, but still converge at less than 0.45 alpha. We were especially gratified to see these results for the strong boundary, for while "too faint" seems a simple perceptual metric, "too strong" does not have an obvious perceptual interpretation. Our results, however, show that while "the fence" is more image and user specific than "too faint," there is sufficient consistency to suggest a perceptual and/or cognitive basis for it.

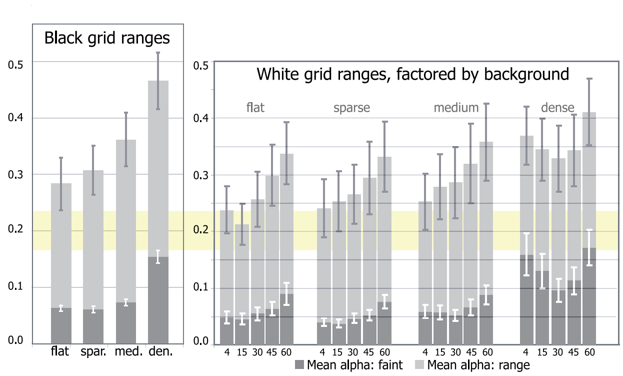


Fig. 18. Faint and range alpha for black and white grid experiments. A range around alpha $= 0.2$ is a good overall setting for these cases.

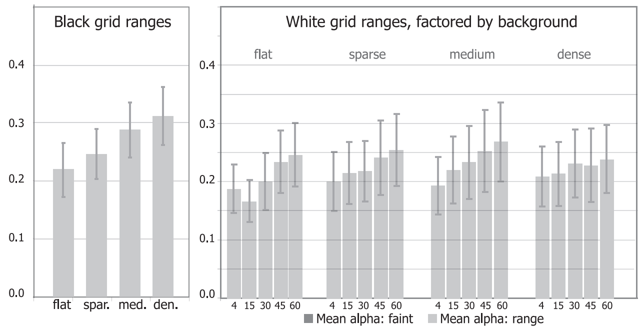


Fig. 19. Range of alpha for the black and white grid experiments.

Another important result is that the background color did not affect where the subjects set the black grid, but was significant for the white grid. This may indicate that the perception of light structures on a dark background is substantively different than dark on light. Or, it may indicate that transparent black grids, which simply darken the colors beneath, are a special case whose properties may not extend to grids of other colors. This is discussed a bit more in the next section.

We were surprised to see the apparently counterintuitive result of grid spacing on the strong boundary; that increasing the grid density decreases the mean alpha settings. We believe this is related to the perception of continuity; rectangular grids are very easy to visually interpolate. This has implications for how much "leeway" may exist for variable grid settings, and warns that these results may not apply to reference structures with different continuity properties.

Fig. 18 summarizes the faint and range alpha values for black and white grids. Fig. 19 shows the ranges alone. While for black grids (Experiment 1) the range becomes larger with density, this is not significant in the white grid (Experiment 2) case. For practical design, the important result is to have a range within which you can set the grid to get an acceptable result. We note that the black grid ranges (Experiments 1 and 3) are moderately higher than the white grid ranges (Experiment 2). A two-factor ANOVA showed this significant especially in the Medium and Dense cases, $p < .01$, again refuting our hypothesis of symmetry between the dark on light and light on dark cases. This shows one very coarse difference that should inform adaptive presentation.

We were interested to see that there is a common, useful range across all experimental conditions within which we can place a subtle grid, highlighted in yellow in Fig. 18. As a practical result, we can recommend using an alpha value of around 0.2 for overlaid grids for images that include a substantial portion of flat backgrounds.

In all of the experiments, the results for the dense plot were significantly different and larger (more salient) than for the other three cases. That the grid needs to be more visually salient against the more visually complex background is not surprising. Understanding how to characterize visual complexity will be the critical issue for this type of analysis.

After the experiments were finished, we analyzed the plots to determine how much of the background was covered, computed as the percentage of totals pixels that were the background color (plus some margin for antialiasing). The results are: 66, 92, 97, and 100 percent (dense, medium, sparse, and flat). In hindsight, if the medium case had been more densely covered, we might have seen a more consistent dependency on density.

9.1 Alpha and Contrast

Luminance contrast is often used to specify legibility thresholds for text and small symbols. It can be specified as a luminance ratio, or in terms of Weber or Michelson contrast. For example, a commonly stated threshold for minimum text legibility is 3:1, which is equivalent to a Michelson contrast of 50 percent. The perception of contrast depends not only on the luminance differences, but the size of the symbol, and whether it is lighter or darker than the background [4].

We computed the contrast between the grid and the background using luminance values computed from the screen calibration and the definition of alpha. We had as an initial hypothesis that our subjects would set their alpha values to maintain a constant contrast with the background, at least for the flat and sparse cases. We also hypothesized that contrast would be a better predictor for the faint boundary than the strong. As is described in more detail in another paper [35], we did not find this to be true, especially for the white grids. The key results from that paper are as follows:

Assuming a simple power relationship to account for the display gamma, we discovered that luminance contrast can be specified in terms of alpha alone for black grids. That is, for transparent black grids, alpha and luminance contrast are equivalent. For example, the Michelson contrast for a black grid rendered with 0.2 alpha is 19.8 percent, for all backgrounds. This suggests black grids may be a special case when using alpha blending.

For the white grid, luminance contrast values are highly dependent on the background, much more so than alpha, suggesting that the appearance of transparency may be more important than contrast. We also note that contrast metrics for light on dark symbols tend to be less consistent than for dark on light.

We also computed the contrast in terms of the color difference value, L^*. We found that ΔL^* was more consistent between the white and black grids than the classic contrast metrics, but was overall more sensitive to background and density than alpha.

9.2 Related Studies

To explore whether trained designers would set these boundaries differently than subjects from the general population, we did an informal study with four subjects self-reported as being trained in design. Each subject did both boundaries for the black and the white grids. We speculated that the designers' data would give more consistent boundaries, and that the boundaries might be fainter. However, we did not find this to be the case. While there are insufficient data to make any statistical argument, in exploring general trends, we found no compelling difference between the way the designers set the boundaries and our subjects from the general population.

Experiment 1 was included as part of a study on the effectiveness of crowd-sourcing for doing perceptual studies by Heer and Bostock [19]. They achieved results similar to ours, except that their boundaries were slightly more salient (larger alpha). Unlike our laboratory setup, they had no control over their display or viewing conditions. Based on the browser-reported User-Agent field, they were able to report a statistically significant difference between users likely to have a 1.8 gamma (older Mac OS systems), and those with a 2.2 gamma for the faint boundary, but not for the strong boundary. This is consistent with what we found in our pilot experiments. Overall, their data support our practical result, that $\mathrm{alpha} = 0.2$ is a safe default value for these kinds of grid overlays.

9.3 Method

A challenge in this research has been how to measure these effects in a way that is both empirically robust and ecologically valid. Perceptual researchers typically favour the traditional psychophysical "staircase" method [11]. However, we elected to use the more exploratory method of having the subject set the desired values for two reasons. In the staircase method, the participant makes yes/no judgments on a variable whose range is successively limited according to each judgment until that range is sufficiently small (i.e., the unit of difference is smaller than the granularity of the measure) to set the final value. This method therefore does not allow a participant to easily go back to a previous setting. Moreover, it moves the task from an interactive one to a passive one. We were interested in how users might explore the different settings of the grid, and even collected data on their explorations, though this has not been analyzed for this paper. We realized that using the mouse as the control tool could introduce bias, but as there was no time constraint, users are well accustomed to mouse interaction, and exploration rather than accuracy was the critical context, this was not a large concern. (We note, however, that the staircase method may indeed be an appropriate approach for some aspects of legibility and discriminability in task contexts in subsequent experiments).

Performance-based measures, on the other hard, allow the evaluation of different configurations in context without the weakness of subject reporting. However, it is difficult to explore all the possible configurations in a reasonable time; this method is more suited to the second phase of our research. We plan to use a performance approach to assess the efficacy of these grid ranges in upcoming experiments.

10 Conclusions and Future Work

We have presented a set of experiments that establish a usable range, defined by alpha, for grids that are neither too faint nor too strong. For the large body of images that are not very dense, a light but useful grid could be created with an alpha value around 0.1, and in all cases, an alpha value of 0.2 falls in the "not bad" range. This is much lighter than the solid black grid ($\mathrm{alpha} = 1.0$) used by default in many visualization systems and technical illustrations. These results reflect professional design advice that transparency is critical to effective layering of elements in an image. We hope that reducing the visual clutter caused by such overly bold grids will be a major benefit of this work.

For these experiments, we wanted to test the hypothesis that for sparse images, contrast with the background would be the dominating effect. Therefore, we designed the cases to cover progressively more of the background, the factor we called "plot density." The consistency of the settings for the flat, sparse, and medium cases suggest that the relationship between the grid and the background is the dominating effect. However, the relationship to luminance contrast is less clear than expected.

The step up in minimum alpha for the dense case, and the general increase in range with density for the black grids illustrates that the visual complexity will be (as expected) a significant factor. However, people do seem to set a reasonably similar specification for the boundaries for the dense image as for the less dense; they are just in a different place. This suggests that if we can characterize the influences, we will continue to find useful metrics.

There remains much to understand about the interplay between overlaid translucent reference structures and the visual complexity of the image they enhance. For example, high-contrast patterns at a spatial frequency similar to that of the grid line interfere strongly, whereas smooth changes in background lightness have minimal effect. A one-pixel wide black grid on a high-frequency black and white noise texture is not visible at any contrast level, but a red one is. It is impossible to imagine algorithmically characterizing all of the cases, but it may be possible to systematically identify solvable, or possibly more usefully, unsolvable ones. Eliminate that which is truly bad and substantial progress will have been made.

We note that the density of the grid itself has a significant if smaller influence on how strong it needs to be, suggesting that as the global "presence" of the grid takes more pixels its local strength can be reduced. We believe this is due to the fact that the eye completes structures from lines. It will be very interesting to examine whether this effect holds true for other reference structures. While the effect was not large enough to be considered as influential in this study, it may have implications for how subtle we can actually make dense grids or reference structures combined with complex images. Further study is indicated.

Our clear practical result is that manipulating alpha works well for ensuring a usable and subtle grid that suits a wide range of situations. For many practical cases, our data suggest that setting a simple level of 0.2 alpha ensures that the grid is both useably visible and comfortably subtle: that it, it sits well below levels judged intrusive. At 0.4 alpha, most users found the grid too strong. This suggests that 0.2 alpha could be considered an example of a JAD for this type of visualization.

We have only demonstrated that this result is valid for a small set of image types and conditions. For a wider variety of images and contexts, more dynamic, adaptive algorithms may well be required. It is clear that we need to develop a more rigorous understanding of what comprises image complexity—the metric we loosely explored as plot density. We used pixel coverage as a coarse metric, but there are many additional factors to be considered, including information complexity, image or task type, and visual and information hierarchies. Especially with respect to the latter issue—critical to how designers think about layering information in a presentation—we were intrigued by the degree to which even small differences in transparency contributed to the sense of grid distance from the underlying image. We think this may have interesting implications for how different kinds of reference structures can be variably emphasized or muted in visualization without requiring explicit user intervention.

In our future work, we want to continue to explore metrics for visual complexity, and their relationship to grid efficacy. Our broader goal is to explore the characteristics of effectively subtle grids and other reference structures over a wide range of images, colors, and tasks, with the hope that we can provide algorithmic approaches to maintaining good design balance in dynamic interactive visualizations.

ACKNOWLEDGMENTS

The authors thank MSc student Billy Cheung for managing and conducting the experiments. The authors thank the reviewers for their comments and suggestions. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] D. Acevedo, C.D. Jackson, F. Drury, and D.H. Laidlaw, "Using Visual Design Experts in Critique-Based Evaluation of 2D Vector Visualization Methods," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 4, pp. 877-884, July-Aug. 2008.

[2] D. Acevedo and D. Laidlaw, "Subjective Quantification of Perceptual Interactions among Some 2D Scientific Visualization Methods," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 5, pp. 1133-1140, Sept.-Oct. 2006.

[3] D. Acevedo, C. Jackson, D.H. Laidlaw, and F. Drury, "Using Visual Design Expertise to Characterize the Effectiveness of 2D Scientific Visualization Methods," *Proc. IEEE Visualization '05, Poster Compendium,* Oct. 2005.

[4] L. Arend NASA Colour Usage website, http://colourusage.arc.nasa.gov, 2010.

[5] A. Bair and D. House, "A Grid with a View: Optimal Texturing for Perception of Layered Surface Shape," *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no 6, pp. 1656-1663, Nov.-Dec. 2007.

[6] E.A. Bier, M.C. Stone, K. Pier, W. Buxton, and T.D. DeRose, "Toolglass and Magic Lenses: The See-Through Interface," *Proc. ACM SIGGRAPH,* pp. 73-80, Aug. 1993.

[7] M. Brill, "Physical Foundations of the Perception Of Achromatic Translucency," RLE Progress Reports No. 117, MIT, 1976.

[8] P. Cavanagh, "Top-Down Processing in Vision," *The MIT Encyclopedia of Cognitive Sciences,* R.A. Wilson and F.C. Keil, eds., pp. 844-845, MIT Press, 1999.

[9] C. Chen, "Top 10 Unsolved Information Visualization Problems," *IEEE Computer Graphics and Applications,* vol. 25, no. 4, pp. 12-16, July/Aug. 2005.

[10] C. Collins, S. Carpendale, and G. Penn, "Visualization of Uncertainty in Lattices to Support Decision-Making," *Proc. Eurographics/IEEE VGTC Symp. Visualization,* May 2007.

[11] T.N. Cornsweet, "The Staircase-Method in Psychophysics," *The Am. J. Psychology,* vol. 75, no. 3, pp. 485-491, 1962.

[12] S. Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten.* Analytics Press, 2004.

[13] J. Findlay and I. Gilchrist, *Active Vision: The Psychology of Looking and Seeing.* Oxford Press, 2003.

[14] G. Fitmaurice, T. Baudel, G. Kurtenbach, and B. Buxton, "A GUI Paradigm Using Tablets, Two-Hands and Transparency," *Proc. ACM Conf. Human Factors in Computing Systems (CHI '97),* pp. 212-218, 1997.

[15] J. Gobell and M. Carrasco, "Attention Alters the Appearance of Spatial Frequency and Gap Size," *J. Am. Psychological Soc.,* vol. 16, no. 8, pp. 644-651, 2005.

[16] E.H. Gombrich, *Art and Illusion: A Study in the Psychology of Pictorial Representation.* Princeton Univ. Press, 2000.

[17] B.L. Harrison and K.J. Vicente, "An Experimental Evaluation of Transparent Menu Usage," *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems: Common Ground,* M.J. Tauber, ed., pp. 391-398, Apr. 1996.

[18] B.L. Harrison, G. Kurtenbach, and K.J. Vicente, "An Experimental Evaluation of Transparent User Interface Tools and Information Content," *Proc. Eighth Ann. ACM Symp. User Interface and Software Technology (UIST '95),* Nov. 1995.

[19] J. Heer and M. Bostock, "Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design," *Proc. Int'l Conf. Human Factors in Computing Systems (CHI),* 2010.

[20] D.H. House, A.S. Bair, and C. Ware, "An Approach to the Perceptual Optimization of Complex Visualizations," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 4, pp. 509-521, July-Aug. 2006.

[21] E. Ishak and S. Feiner, "Free-Space Transparency: Exposing Hidden Content through Unimportant Screen Space," *Proc. Ann. ACM Symp. User Interface and Software Technology (UIST),* conference supplement, Nov. 2003.

[22] C.D. Jackson, D. Acevedo, D.H. Laidlaw, F. Drury, E. Vote, and D. Keefe, "Designer-Critiqued Comparison of 2D Vector Visualization Methods: A Pilot Study," *Proc. ACM SIGGRAPH,* July 2003.

[23] C. Johnson, "Top Scientific Visualization Research Problems," *IEEE Computer Graphics and Applications,* vol. 24, no. 4, pp. 13-17, July/Aug. 2004.

[24] G.E. Legge, G.S. Rubin, and A. Luebker, "Psychophysics of Reading—V. The Role of Contrast in Normal Vision," *Vision Research,* vol. 27, pp. 1165-1171, 1987.

[25] M. Livingstone, *Vision and Art: The Biology of Seeing.* Harry N. Abrams, 2002.

[26] E. Lupton and J.C. Phillips, *Graphic Design: The New Basics.* Princeton Arhictectural Press, 2008.

[27] A. MacEachren and M.-J. Kraak, "Exploratory Cartographic Visualization: Advancing the Agenda," *Computers and Geosciences,* vol. 23, no. 4, pp. 335-343, 1997.

[28] B. MacIntyre and W. Cowan, "A Practical Approach to Calculating Luminance Contrast on a CRT" *ACM Trans. Graphics,* vol. 11, no. 4, pp. 336-347, 1992.

[29] F. Metelli, "The Perception of Transparency," *Scientific Am.,* vol. 230, pp. 90-98, Apr. 1974.

[30] D. Norman, *The Design of Everyday Things.* MIT Press, 1998.

[31] A. Robinsn, "Highlighting Techniques to Support Geovisualization," *Proc. ICA Workshop Geovisualization and Visual Analytics,* June 2005.

[32] H. Senay and E. Ignatius, "A Knowledge-Based System for Visualization Design," *IEEE Computer Graphics and Applications,* vol. 14, no. 6, pp. 36-47, Nov. 1994.

[33] M. Singh and B.L. Anderson, "Towards a Perceptual Theory of Transparency," *Psychological Rev.,* vol. 109, no. 3, pp. 492-519, 2002.

[34] M. Spehr, C. Wallraven, and R. Fleming, "Image Statistics for Clustering Paintings According to Their Visual Appearance," *Proc. Computational Aesthetics in Graphics, Visualization, and Imaging,* O. Deussen and P. Hall, eds., 2009.

[35] M. Stone and L. Bartram, "Alpha, Contrast and the Perception of Visual Metadata," *Proc. 16th IS&T/SID Colour Imaging Conf.,* pp 355-359, Nov. 2008.

[36] M. Stone, *A Field Guide to Digital Color.* A K Peters, 2005.

[37] M. Tory and T. Moller, "Evaluating Visualizations: Do Expert Reviews Work?" *IEEE Computer Graphics and Applications,* vol. 25, no. 5, pp. 8-11, Sept.-Oct. 2005.

[38] E.R. Tufte, *Envisioning Information.* Graphics Press, 1998.

[39] E.R. Tufte, *The Visual Display of Quantitative Information.* Graphics Press, 2001.

[40] C. Ware, *Information Visualization: Perception for Design.* Morgan Kaufman, 2004.

[41] S. Zhai, W. Buxton, and P. Milgram, "The Partial-Occlusion Effect: Utilizing Semitransparency in 3D Human-Computer Interaction," *ACM Trans. Computer-Human Interaction,* vol. 3, no. 3, pp. 254-284, Sept. 1996.

[42] S. Zuffia, C. Brambillab, G. Berettac, and P. Scalaa, "Human Computer Interaction: Legibility and Contrast," *Proc. 14th Int'l Conf. Image Analysis and Processing (ICIAP),* 2007.

Lyn Bartram is an assistant professor in the School of Interactive Art and Technology at Simon Fraser University, a school with the mandate of bringing design, art, science and computing together in technology used by humans. Her background is in applied perception and visualization, and she is particularly interested in understanding why and how certain design and art principles create such effective visualizations and how these might be computationally leveraged in highly dynamic visualization environments.

Maureen C. Stone is an independent consultant working in the areas of digital color, information presentation, interaction and systems. Before founding StoneSoup Consulting, she spent 20 years at the Xerox Palo Alto Research Center where she attained the position of principal scientist. She has a longstanding interest in the use of color in illustration and design, especially interactive tools and systems to support color design and reproduction for both displays and print. As a consultant, she has worked on commercial systems for data visualization and analysis, designing both the color encodings and their user-interface. Her book, *A Field Guide to Digital Color*, was published by A K Peters in 2003, and presents digital color from perception to design. She is an adjunct professor in the School of Interactive Art and Technology at SFU.