The Effect of Colour and Transparency on the Perception of Overlaid Grids

Lyn Bartram, *Member, IEEE*, Billy Cheung, *Student Member, IEEE,* and Maureen C. Stone, *Member, IEEE*

Abstract—Overlaid reference elements need to be sufficiently visible to effectively relate to the underlying information, but not so obtrusive that they clutter the presentation. We seek to create guidelines for presenting such structures through experimental studies to define boundary conditions for visual intrusiveness. We base our work on the practice of designers, who use transparency to integrate overlaid grids with their underlying imagery. Previous work discovered a useful range of alpha values for black or white grids overlayed on scatterplot images rendered in shades of gray over gray backgrounds of different lightness values. This work compares black grids to blue and red ones on different image types of scatterplots and maps. We expected that the coloured grids over grayscale images would be more visually salient than black ones, resulting in lower alpha values. Instead, we found that there was no significant difference between the boundaries set for red and black grids, but that the boundaries for blue grids were set consistently higher (more opaque). As in our previous study, alpha values are affected by image density rather than image type, and are consistently lower than many default settings. These results have implications for the design of subtle reference structures.

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

INTRODUCTION

Visual elements such as grids, labels and contour lines act as reference structures that support the information being presented. They need to be usefully visible – i.e., effective in providing the desired framework or extra data without being obtrusive. Good designers create a layered attention hierarchy that renders such visual metadata so that it can be seen, but only attended to when needed. This work seeks to define computationally tractable metrics and design principles that enable this same layering in dynamic, computer-based visualizations, where the amount and type of information in the image is constantly changing.

Previous work in this area [5][20] looked at black and white rectangular grids overlaid on scatter plot images of different visual density rendered in shades of grey over range of grey backgrounds (light for black grids, dark for white ones). Participants manipulated the transparency of the grids to determine boundary conditions for visual intrusiveness. The results establish a usable range, defined by alpha, for grids that are neither too faint nor too strong. As expected, this range varies with visual complexity, but not widely for the images tested. Similar to grids created by professional designers, the experimental results indicate that a very subtle grid (alpha 0.2) is sufficient for the type of images used in the experiment.

This paper reports on an extension to this work that compares black grids to blue and red ones on a wider range of greyscale image types: scatter plots, two forms of maps (areas and lines), and two abstract images, one designed to maximize interference with the grid. We expected that the coloured grids would be more visually salient than the black ones, resulting in lower alpha values. Instead, we found there was no significant difference between black and red, but that the boundaries for the blue grids were set consistently higher (more opaque). The results for the black grids are consistent with the previous experiment and contrary to our expectations, do not vary significantly across image type except for a "worst case" abstract image. As in the previous experiment, the results do depend on image complexity.

This work is part of a larger project to create metrics for effective visual design based on a study of vision, perception and design practice. The overall goal of our research is to understand and quantify subtle aspects of visual representation such that they can be algorithmically manipulated to match human requirements in interactive and dynamic conditions.

1 RELATED WORK

Designers create unobtrusive reference structures by varying visual parameters, including colour, contrast and transparency to manipulate the Gestalt principles of figure and ground [25]. The overall goal is to create visual layers, where whatever constitutes a "figure" is well defined with respect the "ground." Grids and other visual metadata live somewhere between these layers, where they are visible only when attended to. The art psychologist Gombrich describes a visual *middle ground* [10] where features can be "extruded" into the foreground or "receded" into the background by slightly changing the degree of attention. Recent research in task-directed vision and attentional effects on visual acuity promise some perceptual and cognitive ground for these effects [8][9]. Recent research on managing attention in graphic displays emphasizes the importance of reducing clutter using techniques like layering and, more broadly, managing attention when creating effective visual displays [17].

The grids used in these experiments are integrated into the underlying image using transparency (alpha blending). This makes the grid adapt to the colour of the underlying image, and keeps it from entirely obscuring any of the underlying information. Transparency has been used in visualization as a representation dimension, to show uncertainty by making uncertain objects less opaque [6], overlaying a transparent wash for highlighting [16] and more generally for reducing screen space limitations by overlaying objects or features. MacEachren includes it in a list of cartographic visual variables [14]. In the user-interface domain, Harrison [11][12] 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). New display technologies that physically incorporate transparent layers have generated recent interest in studying transparent layers in the presentation of information [25].

This work integrates techniques from design and perception to develop methods and metrics for creating effective visualizations. Recent studies by Acevedo et al. [1][2] and Tory and Möller [22]

- *Lyn Bartram is with Simon Fraser University, E-Mail: lyn@sfu.ca.*
- *Billy Cheung is with Simon Fraser University, E-Mail: bcc7@sfu.ca.*
- *Maureen Stone is with StoneSoup Consulting, and also with Simon Fraser University, E-Mail: stone@stonesc.com.*

Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

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. Recent studies of the perception of scatter plots [18] use psychophysical techniques to evaluate their effectiveness, and to create metrics for optimizing their presentation.

The work in this paper builds directly on previous studies of overlaid rectangular grids [5][20] that reported a set of experiments in which the subjects define a pair of boundaries that define a range of visible yet subtle legibility. This work introduced a new metric called the *JAD*, or *Just Attendable Difference*. A JAD is similar to the JND (just noticeable difference) 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. The results indicate that an effective grid can be created for the images used in the experiment by setting alpha to 0.2, and offer this as an example of a JAD for overlaid grids. These experiments tested the effect of visual complexity, background colour and grid spacing on the resulting boundary values, using a black or white grid overlaid on scatterplots of varying densities (similar to the bottom pair of images in Figure 1). Only visual complexity was found to be significantly important. The work in this paper extends the support for the JAD to a wider range of image types and introduces the effect of colour.

Bostock and Heer used crowd sourcing to replicate our earlier greyscale experiments and found similar effects [13]. They used the same images and the same procedures, but were naturally unable to provide the same controlled viewing conditions we used in our laboratory. However, their results also concluded that a value of 0.2 for alpha would provide good results, even in these less controlled viewing environments.

2 EXPERIMENT

Rather than seeking some ideal grid setting, the approach in this work is to define two boundaries that bracket a range of grids that are sufficiently salient to be usable, but not so strong as to be obtrusive. The *faint boundary* defines the faintest useable grid, and the *strong boundary* describes where the grid is too strong, creating visual clutter, or becoming “a fence” in front of the background, rather than being integrated with it. (We note the “fence” terminology came directly from participants in our earlier studies describing the effect of the grid appearing to “pull away from” and “come in front” of the image when it became too intrusive [5].) In this experiment, the goal was to explore the effect of colour and image type on the faint and strong grid boundaries.

As in the previous experiments, participants were asked to adjust the alpha value of a grid with a constant line weight of one pixel (about 1.5 minutes of arc) and a constant colour (black, red, or blue) over a set of four images at two density levels. The images were chosen to represent different common forms of visualizations used with rectangular grids (scatter plots and two forms of maps), plus a pair of abstract images representing extreme density conditions. Two types of data measures were collected: 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.)

A 2 (image density) x 4 (image type) x 3 (grid colour) factorial design yielded 24 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 3 repetitions of 8 images overlaid by each of the 3 grid colours, resulting in 72 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.

Fig. 1. The images used in the experiment, in pairs with the dense image on the left, the sparse on the right. From the top: Abstract, Area map, Line map, and Scatter plot.

2.1 Image types

The images used in the experiment are shown in Figure 1. Two levels of image density (Sparse and Dense) were created for each of four image types: Scatter plots (Plot), Line maps (Line), Area maps (Area), and Abstract.

The Plot images represent the type of plots used in the previous study, and generally represent data plots and graphs. The dense plot covers the background more thickly with dots than the sparse case, which shows primarily the background colour. The scatter plot images were generated using Tableau Software’s desktop system, using synthetically generated data.

The Line maps were created from contour maps and were chosen to see how an image that is primarily lines interacts with the lines in the grid. Visually, they consisted of a background, mostly of a constant colour with some added features like rivers and lakes, and contour lines on top of the background. Colour lines (shades of grey) and line thickness were varied. For the sparse case, two levels of thickness were used to visualize one data type (elevation). A thicker contour line was used for every 5 small steps in the change of elevation. A relatively flat area (in our case, a glacier) was chosen to obtain a low density map. For the dense map, an area with more change in elevation was chosen. There were also lines to represent rivers that only loosely visually matched with the change of elevation (with cross over between two types of lines).

The Area maps were taken from Open Street Map street maps. (www.openstreetmap.org). visually, they contained mostly thick lines (as streets) and shapes (e.g. a park) that filled up the 2D plane with shades of grey. In contrast to the plots and the contour map, there was no fixed background colour. The black text labelling locations on the map was removed to avoid introducing effects having to do with the interpretation of text. For the sparse case, a remote town in the UK was used. For the dense case, we used a section of a downtown London map.

The Abstract cases were chosen to create extreme examples of background complexity with respect to perception. In our previous study where the images were displayed on different background colours (shades of gray) we had included images of a simple field of the background, as an "extreme sparse" case. We had also experimented with various types of extremely dense images, to indicate the other extreme, but did not include any in the experiments. For this experiment, we repeated using a flat coloured area (Abstract Sparse) For the Abstract Dense case we chose an image filled with black-and-white noise at a spatial frequency similar to the grid (one pixel). Informal experiments had indicated that this kind of texture interfered strongly with the perception of black grid lines without physically obscuring them (like large solid black areas would). We expected that adding colour to the grid would help significantly in separating it from the background noise. We included the flat field for comparison to the previous experimental results.

2.2 Grid colour

The perception of colour is complex, so there are many factors that could influence the perception of a coloured grid. For this experiment, we needed a limited number of colours to keep the experimental permutations tractable. We wanted to include black for comparison with the previous study, and to see the influence of image type on the black grid settings. To provide some ecological validity, we chose red and blue, which are commonly used grid colours in maps. We picked colours that looked like the grid colours of well designed maps and matched them with respect to perceived brightness: Blue RGB = (36, 104, 217), Red RGB = (210, 9, 4), L*=45, but did not apply any other metrics to their selection. Our expectation was that the significant difference would be colour vs. black so the specific colour values were not especially critical. This turned out not to be true, with the blue grid perceived differently than the red or the black. The factors that might contribute to this result will be addressed in the Discussion section.

2.3 Method

We used the same experimental method from a previous study in which we investigated factors of image density, background and grid spacing [5]. Each participant was asked to perform two tasks, one for each boundary condition. One was to specify the point where "the grid is just perceptible without being **unnoticeable** or **unusable**." (Faint grid).

Fig. 2. One of the training images with a faint blue grid, showing the experimental setup.

The other was to adjust the grid "to meet your best judgment of how obvious it can be before becoming too **intrusive**" (Strong grid). The instructions stated that there were no time restrictions or correct answer, and that participants were to play with the settings until satisfied with the result. We note that this task does not measure the effectiveness of the grid with respect to image performance (how much it supports or distracts from the underlying information): these are more difficult data to elicit. Rather, we took a "designerly" approach, asking people for their best judgment of these settings, similar to previous work by Acevedo and Laidlaw [1].

Using a standard LCD computer monitor, 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). 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). Once the participant was satisfied with the result, (s)he would press the Next button to proceed to the next image. Participants could practice on a set of training images for an unlimited time. Most of the users performed the task on the same Apple Cinema display, but a few used an equivalent display in another lab, for convenience. The displays were visually calibrated for consistent brightness and contrast. Figure 2 shows one of the test images with an overlaid red grid displayed in the application driving the experiment. To ensure that the subjects did not accidently go to the next image without adjusting the grid, the Next button did not appear until at least three adjustments were made.

As in our previous studies [5], we set the initial alpha value at 0; that is, when the trial began there was no apparent grid on the image. By increasing opacity participants "added" it to the image.

2.4 Hypotheses

H1. Grid colour would affect alpha setting. We expect the grid with colour (Red and Blue) would result in a lower alpha setting than the Black grid, as hue contributes an additional cue to the segregation between the greyscale data layer and the colour reference structure.

H2. Alpha would be affected by Density, with Dense images requiring higher alpha settings than Sparse ones. At the extreme, the Gaussian noise plot (Abstract Dense) would be a pathological case requiring much stronger values than the more realistic images.

H3. Alpha settings would vary with Image Type. We were especially interested in the Line map, because it uses the same visual element (thin line) as the grid, varying only in curved vs. straight. We would expect that its settings would be significantly different than the Plots or the Area map.

H4. There would be more variability in the Strong settings than the Faint ones. The Faint boundary is akin to the psychophysical property of minimum visibility. The Strong setting is more subjective.

H5. Image Type and Density will affect the Range of the alpha. The Range of alpha is the difference between the Faint and Strong grid. More specifically, we expect a higher range for data structure with higher density, to address the visual interference in the dense image.

2.5 Participants

15 participants, roughly distributed by gender, with normal or corrected-to-normal vision were paid to undertake the experiment. Participants were primarily students from design, computing and business backgrounds.

3 Results

We analysed the effects of Image Type, Density and Grid Colour on three different sets of data: the Faint alpha settings, the Strong alpha settings, and the Range (the difference between means of the two). Image Type was significant in all three sets of data, but due only to

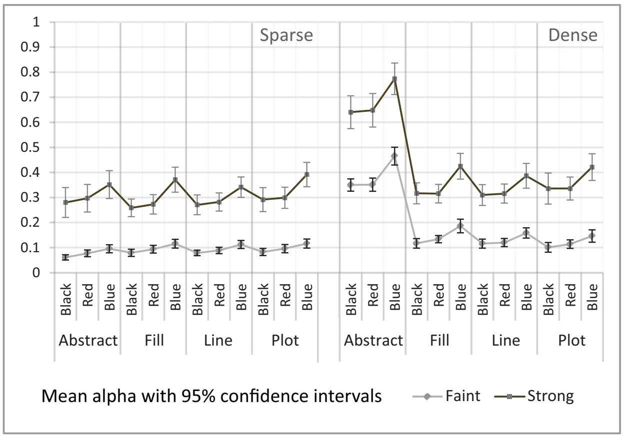
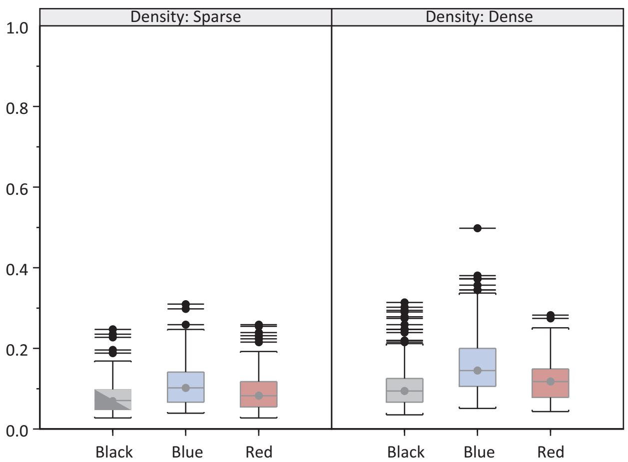


Fig. 3. Mean alpha values for each image for each colour.

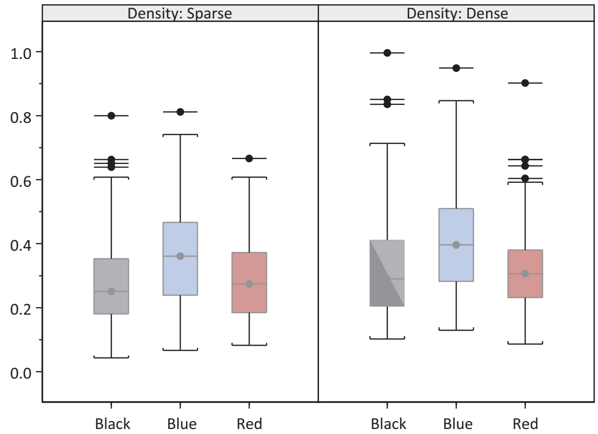
the influence of the Abstract images. When we removed these images, Image Type no longer had a significant effect, refuting H3. The Line maps were no different than the Area maps or the Plots. Grid Colour was also significant, though not in the way we expected: the Red and Black grids were quite similar, but the Blue grid used stronger values. Figure 3 shows the mean alpha set for the dense and sparse cases for each image type for each colour. The unusual results for the Abstract Dense image is quite clear, as are the stronger values needed for the Blue grid in all cases (even the pathological one). The interesting result for visualization is that there is no difference among the images that simulate reasonable use cases. Therefore, the Abstract case has been removed from most of the results discussed in the rest of this section.

A two-way ANOVA showed both Density (F(1,1074)=2.98, p<.001) and Grid Colour (F(2,1074)=.26, p<.001) to be highly significant for the Faint data. There was a very small interaction between them (F(2,1074)=.03, p<.023), caused by the Abstract Dense image. The interaction disappeared when the Abstract images were removed. In the Strong data, we saw significant effects for Density (F(1,1078)=109.9, p<.001) and Grid Colour (F(2,1077)=24.03, p<.001). As expected, the Abstract Dense image (Gaussian noise texture) proved to be pathological. To focus on more realistic images, we have removed the Abstract image type from the remainder of the results presented in this section. Table 1 shows all alpha means in the remaining data set: the lowest mean value is the Faint Black Sparse setting at 0.08; the highest is the Strong Blue Dense value at 0.41.



Distributions of alpha values for the Faint task, by plot type and by colour

Fig. 4a. Alpha factored by colour for the Faint grid setting, for Sparse and Dense tasks (abstract images removed).



Distributions of Alpha values for the Strong task, by plot type and by colour

Fig. 4b. Same as Figure 4a, but for the Strong setting.

Table 1. Alpha means by Task, Density and Grid Colour over all Image Types, Abstract images excluded.

Task	Density	Black	Red	Blue
FAINT	Sparse	.08	0.093	0.114
	Dense	.111	0.123	0.164
STRONG	Sparse	0.274	0.285	0.368
	Dense	0.321	0.322	0.411

There were no interactions between any of the factors in the Strong data. As predicted (H2), a one-way ANOVA showed Density had a significant effect on both the Faint and Strong alphas: F(1,808)=70.4, p<.001 (Faint), F(1,808) = 15.3, p <.001 (Strong). However, the size of this effect was less than we expected: approximately 0.05 alpha in both Faint and Strong grids. There was little variation in the Faint settings overall, and as expected more variation in the Strong, confirming H4.

3.1 Grid colour

As hypothesized (H1) Grid Colour proved highly significant in both the Faint (F(2,807)=35.3, p<.0001) and Strong(F(2,807)=31.67, p<.0001) cases. To our surprise, this effect proved to be due to the difference between the Blue grid and the other 2 colours, Black and Red, rather than the difference between a black and a coloured grid. The mean alpha for Blue is consistently higher for both the Faint and Strong cases. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the Faint Blue condition (M = 0.15) was significantly different than the Faint Black (M = 0.09) or Faint Red (M=0.1) conditions. In the Strong case, Blue (M=0.39) was significantly higher than Red (M=0.3) or Black (M=0.3). In fact the difference between the Blue grid and the other colours was the most pronounced effect of all the factors. There were no interactions between Grid Colour and other factors. Figures 4a and 4b show the mean alpha values for each Colour for each Density for the Faint and Strong boundaries, respectively. The effect of grid colour was also pronounced on the Abstract images in both the Faint (F(2,267)=4.94, p<.007) and the Strong (F(2,267)=3.64, p< 0.027) tasks. This was again due only to the effect of the Blue grid. The red grid was not set significantly differently from the black one.

3.2 Range data

We then turned our attention to the Range data (Figure 5). Each participant saw three replications in total of each condition in the Faint and Strong tasks; we took the mean of each condition per subject and took the difference between them to calculate the Range. We found no effect of Image Type for Range. To our surprise, Density also had no significant effect – that is, people were fairly consistent

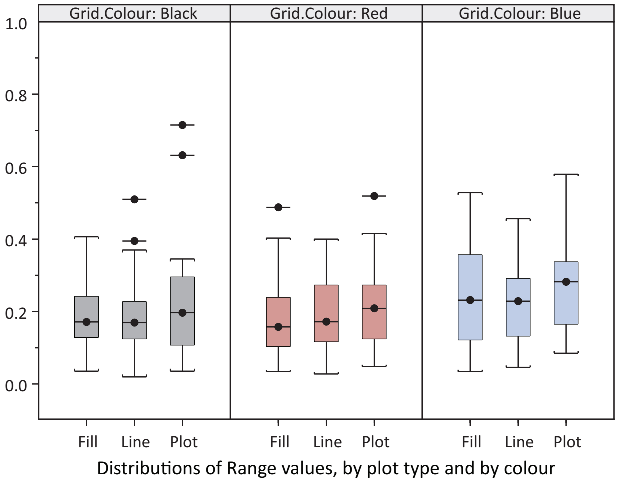


Fig. 5. Range factored by plot type and by colour (abstract images removed).

in their differences between Faint and Strong settings across density. Thus their settings for Strong grids were largely dependent on their "starting point" of the Faint setting. However, we found Grid Colour to be significant ($F(2,267)=5.44$, $p<.005$.) Again this was due only to the influence of the Blue grid. Not only did people set its Faint setting higher, they consistently increased it more than the other 2 colours for the Strong setting.

This measure of range in which we take the difference of averages can be problematic as we cannot make the assumption that each of the distributions (Faint and Strong) is equal (the difference between subjects shown in Figure 6 illustrates this). While there is overlap in these noisy data (Figure 5 shows the 95% confidence intervals), however, it is still apparent that the Blue grid has a wider range.

3.3 Summary

Figure 6 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 is plotted by a line. This lets us compare the relative sizes of the effect on means for different factors. The spread in mean settings is greatest across subjects, skewed by two outliers.

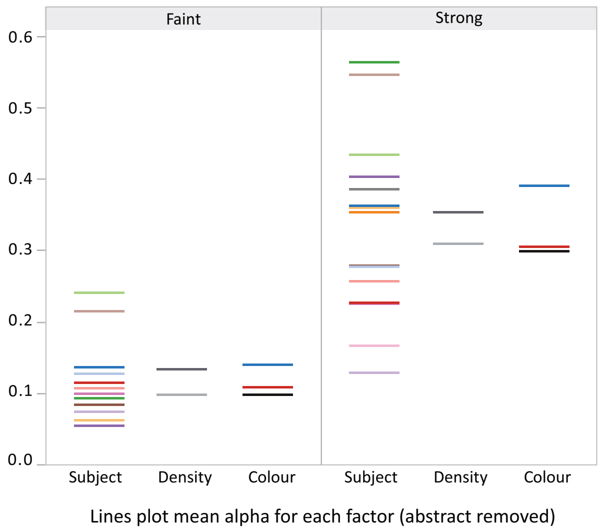


Fig. 6. Average alpha as function of Subject, Density and Colour for each task. In the Subject column, the colour indicates individual participants. In the Density column, it marks the Dense (dark) vs. the Sparse images. In the Colour column, it indicates the colour of the grid.

This variability is most pronounced in the Strong grid. There is a marked range between the Faint and Strong means for each participant. There is a small consistent difference in Density independent of the grid task. The higher value for the Blue grid can be clearly seen.

4 Discussion

In this study, we present two new results. The first is that colour is important, but that the effect is not simple to predict: red did not behave differently than black, and blue was significantly less salient than red. The second is that image type had no significant effect for all the images that simulate reasonable use cases.

We can only speculate at this point why red and blue behaved differently, and why there was no significant difference between red and black. Given how effectively a coloured object "pops out" of a grey background, we expected adding colour to make the grids significantly more visible, resulting in lower alpha values. For the black grid, transparency affects only the contrast in lightness with respect to the background. However, any coloured grids will initially be lighter (higher L*) than the black grid. Perhaps the additional "colour contrast" of the red grid just balanced out the difference in lightness contrast. As well as hue and lightness, colours can be compared by their colourfulness or chroma. The two colours matched with respect to L*, but the red colour is significantly higher chroma (measured as C_{ab} in the CIELAB space) than the blue one. Since chroma is also reduced as transparency increases, the difference could be critical.

Another possibility for the difference between red and blue is the special perception of blue, which is influenced by the eye's relative insensitivity to short wavelengths and the limited spatial acuity of the short wavelength cones. While this effect is very strong for the extremely pure blue primary of a CRT display (less so for LCD displays), the blue used in this experiment is quite different, containing significant amounts of green and red. Informally exploring different colours of red and blue balanced for both lightness and chroma suggest that it really is the hue difference that matters. Designers have observed [15] warmer colours (red, orange) seem visually more salient than cool colours (green and blue). We clearly have insufficient data to support this in any significant way, but that our results match this observation is intriguing.

The second result is that – again contrary to our expectation – image type, at least for the classes of images we selected, seems unimportant. This result needs to be tempered by the observation that while the images we chose were representative of maps and plots – very common visualizations – they were not actually completely realistic. Labels, icons and other highly salient features had been removed, as we were primarily interested in the potential interactions of lines and filled areas with single line grids. It remains to be tested if completely realistic images would assure the same results. Nonetheless, the fact that image type seems less important than we anticipated is interesting because it suggests that these results will generalise across a variety of contexts. There are many dimensions of image complexity that remain to be explored, and image coverage (Density) is only the first simple one. That it plays a part in balancing the grid with the image, but that the elements of that density that differentiate different image types (such as lines, filled areas and dots) do not indicates we need more informative models of how these layers work together.

In contrast, the Abstract Dense case clearly shows that some forms of visual complexity will have a significant impact. We might speculate that overlaid grids are only commonly used on the types of images where they easily perceivable. Design practice has, in effect, created the class of images where such subtle grids are easy to use. Perceptually speaking, gridlines are easiest to see on a low-frequency background of sufficient luminance contrast; the Abstract Dense case was designed to have no such areas, and therefore the grid is much harder to see.

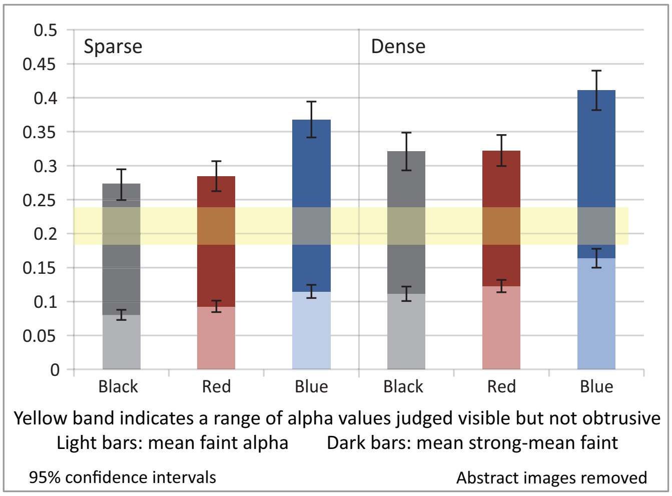


Fig. 7. Usable range based on these experiments

Our results for this experiment are consistent with our previous studies, in which a 0.2 alpha setting fell within the boundaries across all image conditions in both dark and light grids, though for blue grids a slightly higher value would be more robust. (see Figure 7)

In terms of our JAD metric, and given the ranges we saw across conditions, we surmise that there are a few practical JAD levels: *not there* (unusably imperceptible), barely *there* (usable), *comfortably there* (usable), and *too strong* (intrusive). Depending on the grid and image density, the range between barely there and comfortably there may expand or shrink, but as long as those boundaries are effectively respected, intermediate settings will serve across a wide variety of images. Our experiments continue to show a "safe" JAD for overlaid grids at alpha 0.2.

There are, however, a number of questions around how well these results will extend beyond the limited application of these images and task. One relates to the starting alpha of the grid. We have run a pilot experiment using the same experimental design with a different starting alpha of 0.5. The intention was simply to provide a visible, rather than an invisible starting position for the grid, with half-opacity somewhat arbitrarily chosen so that the grid colour was visible. We found the results significantly different than when the experiment started with alpha at 0, especially for the strong case. In neither case did we find any significance of trial order, refuting the idea that we are creating a learning effect by starting at 0.

Inspection of how participants manipulated the grids suggested that they took the half opacity as a minimum starting point. While this is still only a pilot, the results suggest that what people define as the "fence", or the upper boundary, is strongly influenced by what they first see of the grid. The start value of 0.5 alpha was equal to or higher than many Strong settings in the start 0 study. When we asked people to set the grid to the Strong level starting from alpha 0 (effectively no grid) they set it much lower. We conjecture that that perhaps the starting point of the half opaque grid was already "strong" enough, and increasing it made it more obvious but did not materially change its effect, whereas the incremental changes from "no grid" to a "strong enough" in the lower half of alpha were noticeable enough to produce the effect of the grid moving in front of the image. However, clearly this requires further investigation

A second question relates to how well the desired settings according to people's judgment will actually affect performance when using the visualization. We will need to determine how best to measure utility and performance with respect to what intrusiveness impacts, as such studies represent the next step in validating how to automatically adapt how layers of semi-transparent information can be useably and appealingly composited.

5 Conclusion and Future Work

Grids may seem to be a limited case, but they have interesting properties for the more general class of overlaid reference structures: they are ubiquitous, they cannot be too intrusive without interfering with the image content, and they have a limited set of representational options (thin lines, colour and labels) that constrain legibility. We believe our results are promising for the larger issues of subtly layering information such as reference structures onto and into complex visualizations.

Borrowing from artists and designers who carefully manipulate visual hierarchies, we are exploring a yet incompletely defined metric of "just attendable difference" to quantify layered reference structures. It is this notion of attendability 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 visualisation. We believe that *attendable* (JAD) is bigger than simply *perceptible* (JND), but quantifying it in an applied context is more challenging. Rather than directly attempt to measure it, we use an empirical method that draws equally from perceptual methods on the one hand and design practice of careful iterative manipulation on the other to determine levels of legibility, aesthetics and utility. This combination is core to the emerging area of computational aesthetics in visualization.

These experiments clearly point to further work on the perception of grids. These include different colour characteristics and the introduction of further complexity into the backgrounds, such as labels and symbols, to better simulate realistic imagery. We need to explore grids on coloured backgrounds. We would also like to develop more task oriented experiments, though this will require a significantly different methodology.

We would also like to expand our efforts to a larger set of reference structures, including labels, user interface markers, menus, contour lines and general transparent overlay regions such as washes and highlighting. We will expand the conditions under which we examine these. As well, we are exploring the idea that we can manipulate separability and distance between layers with other techniques than transparency like brightness and motion, commonly used in interactive environments such as games [19]. Finally, we intend to move to more performance- and task-oriented measures for measuring intrusiveness and legibility in overlays.

Acknowledgments

This work was supported in part by a grant from the Canadian GRAND Network of Centres of Excellence. The second author would like to thank Athabasca University for supporting his study leave at Simon Fraser University.

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 Transactions on Visualization and Computer Graphics*, vol 14, no. 4, pp. 877–884, Jul.-Aug. 2008.

[2] D. Acevedo and D. Laidlaw, "Subjective Quantification of Perceptual Interactions among some 2D Scientific Visualization Methods," *IEEE Transactions on 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," *IEEE Visualization'05*, Poster Compendium, October 2005.

[4] A. Bair and D. House, "A Grid with a View: Optimal Texturing for Perception of Layered Surface Shape," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1656–1663, 2008.

[5] L. Bartram and M. Stone, "Whisper, Don't Scream: Grids and Transparency," *IEEE Transactions on Visualization and Computer Graphics*, in press, 2011.

[6] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose, "Toolglass and Magic Lenses: The See-through Interface," Proceedings of SIGGRAPH '93 (Anaheim, CA, August 1-6, 1993), in *Computer Graphics* Proceedings, Annual Conference Series, pages 409–412. ACM SIGGRAPH, ACM Press, July 1993.

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

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

[9] J. Gobell and M. Carrasco, "Attention Alters the Appearance of Spatial Frequency and Gap Size," *Journal of the American Psychological Society,* vol. 16, no. 8, 2005.

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

[11] B. L. Harrison and K. J. Vicente, "An Experimental Evaluation of Transparent Menu Usage," *Proc. of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground*, M. J. Tauber (Ed.), pp. 391–398, April 1996.

[12] B. L. Harrison, G. Kurtenbach, and K. J. Vicente, "An Experimental Evaluation of Transparent User Interface Tools and Information Content," *Proc. of the 8th Annual ACM Symposium on User Interface and Software Technology*, pp. 81–90, 1995.

[13] J. Heer and M. Bostock, "Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design," *Proc. of the 28th international conference on Human factors in computing systems (CHI '10)*, pp. 203–212, 2010.

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

[15] P. Meggs and A. Purvis, *History of Graphic Design*, John Wiley & Sons, Hoboken, NJ, 2005.

[16] A. Robinson, "Highlighting Techniques to Support Geovisualization," *Proc. of the ICA Workshop on Geovisualization and Visual Analytics*, 2005.

[17] R. A. Rensink, "The Management of Visual Attention in Graphic Displays," *Human Attention in Digital Environments*, Claudia Roda (Ed.), Cambridge University Press, Cambridge, England, 2011.

[18] R.A. Rensink and G. Baldridge, "The Perception of Correlation in Scatterplots," *Computer Graphics Forum*, vol. 29, no. 3, pp. 1203–1210, June 2010.

[19] M. Seif el-Nasr, L. Bartram, D. Milam, and D. Moura, "Visual Composition and Layering in Video Games," *Foundations of Digital Games 2011*, submitted for publication.

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

[21] M. Stone, *A Field Guide to Digital Colour*, A. K. Peters, Boston, MA, 2005.

[22] 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.

[23] E. R. Tufte, *Envisioning Information*, Graphics Press, Cheshire, Connecticut, 1998.

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

[25] C. Ware, *Information Visualization: Perception for Design*. Morgan Kaufman, San Francisco, CA, 2004.

[26] B. L. W. Wong, R. Joyekurun, H. Mansour, P. Amaldi, A. Nees, and R. Villanueva, "Depth, Layering and Transparency: Developing Design Techniques," *Proc. the 17th Australia conference on Computer-Human Interaction (OZCHI '05),* 2005.