Originally published: Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, 1993, pp. 343–347. **Republished**: G. Perlman, G. K. Green, & M. S. Wogalter (Eds.) *Human Factors Perspectives on Human-Computer Interaction: Selections from Proceedings of Human Factors and Ergonomics Society Annual Meetings*, 1983–1994, Santa Monica, CA: HFES, 1995, pp. 321-325.

COLOR VERSUS TEXTURE CODING TO IMPROVE VISUAL SEARCH PERFORMANCE

Gary Perlman and **J. Edward Swan, II**
Department of Computer and Information Science
Ohio State University, Columbus, OH, 43210 USA
{perlman,swan}@cis.ohio-state.edu

An experiment is reported in which the relative effectiveness of color coding, texture coding, and no coding of target borders to speed visual search is determined. The following independent variables were crossed in a within-subjects factorial design: Color coding (present or not), Texture coding (present or not), Distance between similarly coded targets (near or far), Group size of similarly coded targets (1, 2, 3, or 4), and a Replication factor of target Border width (10, 20, or 30 pixels). Search times, errors, and subjective rankings of the coding methods were recorded. Results showed that color coding improved search time compared to no coding, but that texture coding was not effectively used by subjects, resulting in nearly identical times to uncoded targets. Subjective preference rankings reflected the time data. The adequate power of the experiment along with the results of preparatory pilot studies lead us to the conclusion that texture coding is not an effective coding method for improving visual search time.

Keywords: Coding, color; Coding, texture; Color; Design; Empirical studies; Evaluation, subjective; Highlighting; Models and theories; Screen output; Visual search

1 INTRODUCTION

Coding is a pervasive technique for identifying attributes of objects. Of the 944 MITRE user interface guidelines (Smith & Mosier, 1986), 61 are devoted to coding methods including color-, tonal-, blink-, shape-, line-, and typographical-coding. The popularity of business graphics and black-and-white laser printing has made another type of coding more popular: fill patterns of different textures. Two of the 61 above guidelines on coding mention the use of textured fills, but these are only at the level of suggestions that simple fills might be used; there is no empirical basis provided for the use of the method. Ware and Knight (1992) demonstrate differential discriminability of textures, but not their effectiveness for improving performance. We know of no empirical studies establishing the efficacy of texture coding.

The most popular coding method to investigate has been color coding, and it has been shown to be highly effective in many studies (Helander, 1987), and guidelines on the use of color are popular (e.g., Rice, 1991). As the evidence for the effective use of color has increased, the use of color computer displays has also gained popularity, but there is a large proportion of computers (many workstations, but also Macintosh and PC's — particularly portables), that present a monochromatic view of the world. Also, almost no printers present colors. The result is that texture (fill patterns such as hatches and diagonal lines) is a popular form of coding. Tufte (1983) offers advice to avoid fills that create visual vibration, advice that is echoed by Smith and Mosier (1986) and Galitz (1993). Still, even with low vibration textures, the question arises whether textures can be an effective coding method.

Our initial hypothesis was that texture coding could be of intermediate effectiveness, better than no coding, but not as effective as color. This would allow texture coding to be used when color was not available, or when it was advantageous to use color to code other attributes. Our immediate application was that of a multi-window environment in which many windows from many applications were present. In an interactive environment in which windows compete for limited space, the casual use of a system in which one application can open many windows often results in a clutter of windows of different types. One solution is for users to become better organized, using a multiple-application structuring system like Rooms (Card and Henderson, 1987), but a variety of window applications can still appear in a room. If applications were coded, then finding a target window would be facilitated.

We devised an experiment to determine if texture coding could help users find windows in well-organized and poorly organized displays. We included color coding and no coding in the experiment so that we could measure the <u>relative</u> effectiveness of texture coding. We included a combined color+texture condition because redundant codes have been found to be significantly better than individual codes (Swierenga et aI, 1991).

Originally published: Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, 1993, pp. 343–347. **Republished**: G. Perlman, G. K. Green, & M. S. Wogalter (Eds.) *Human Factors Perspectives on Human-Computer Interaction: Selections from Proceedings of Human Factors and Ergonomics Society Annual Meetings*, 1983–1994, Santa Monica, CA: HFES, 1995, pp. 321-325.

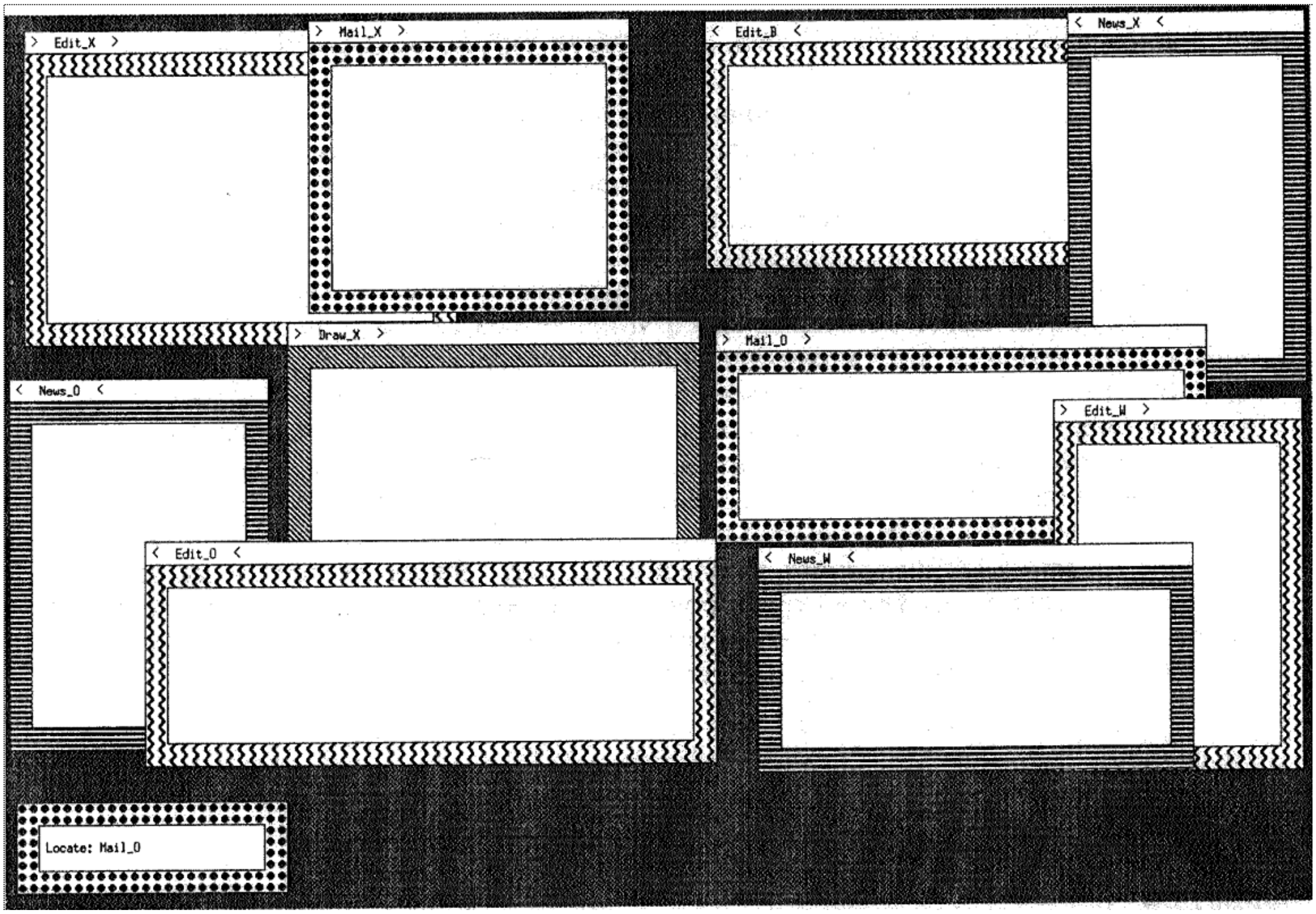

Figure 1. Textured window configuration with *far* windows with *medium* (20-pixel) borders. The target (Mail_O) is displayed at the lower left, for which the correct response would be ">" .

2 METHOD

2.1 Experimental Design

2.1.1 Independent Variables

Color and Texture Coding: Two primary independent variables were Color (either present or not) and Texture (either present or not) crossed in a 2 x 2 within-subject matrix. Four colors were used: red, green, cyan, and yellow; chosen based on recommendations of Rice (1991), with the saturated blue being replaced by the brighter cyan. Four textures were chosen (see Figure 1) to provide moderate variation of overall density, spatial frequency, and similarity of description. When Color and Texture coding were both present, the black texture was superimposed on the color borders, producing a dimmer color. The order of presentation of the 4 Coding conditions in the 2 x 2 matrix was counterbalanced with a between-subjects latin square.

Trial / Group Size: There were always 10 windows on the screen, enough so that the locations of windows could not be memorized easily and not so many that the task was unrealistic for a large-screen workstation. Each window was the target in a block of 10 trials; the order of targets was a permutation randomly generated for each block, independent of any other blocks. Four Colors were used for the 10 window borders: red, green, cyan, and yellow; 4 windows were one color, 3 another, 2 another, and 1 the last (the assignment of colors to number of windows was counterbalanced in a between-subjects latin square). Four Textures were used for the 10 window borders (see Figure 1), 4 windows were one texture, 3 another, 2 another, and 1 the last (the assignment of textures to number of windows was counterbalanced in a between-subjects latin square). The combination of the 4 colors and 4 textures were counterbalanced in a between-subjects latin square.

Originally published: Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, 1993, pp. 343–347. **Republished**: G. Perlman, G. K. Green, & M. S. Wogalter (Eds.) *Human Factors Perspectives on Human-Computer Interaction: Selections from Proceedings of Human Factors and Ergonomics Society Annual Meetings*, 1983–1994, Santa Monica, CA: HFES, 1995, pp. 321-325.

Distance between Similar Windows: Each subject saw two window Distance configurations: a Near configuration, in which similar windows (Color and/or Texture) were adjacent on the screen, and a Far configuration (shown in Figure 1), in which similar windows were distributed all over the screen. A pilot study showed no differences between one-sized tiled and variable-sized slightly-overlapping windows, so only the more natural (latter) configuration was used, which roughly controlled for amount of visible border (see Figure 1). Although a pilot study showed no differences for different similar distance configurations, two Near configurations and two Far configurations were counterbalanced in an incomplete between-subjects latin square for increased generality of results. Within the 4 Coding conditions, half the subjects saw the Near configuration followed by the Far configuration, and half saw the opposite order.

Replication / Border Width: To allow users to gain experience with the combination of window Coding and Distance, there were 3 replications of each block of 10 trials. To explore an additional factor between these replications, 3 Border widths (small = 10 pixels, medium = 20 pixels (shown in Figure 1), and large = 30 pixels) were presented to each subject. The same border Coding and window Distance was used for the 3 Replications / Border widths, and in no case was the border so small that the texture was unidentifiable, because each Texture was designed on a 10 x 10 pixel grid. Half the subjects saw only increasing Border widths, and half saw only decreasing Border widths. (Completely counterbalancing the Border widths was thought to be unnecessary because the variable is on at least an ordinal scale.)

Presentation order of conditions maximized the continuous practice subjects had with each (Color x Texture) coding scheme. The within-subjects Color (2) x Texture (2) x Distance (2) x Trial within block (10) x Border width / Replication (3) resulted in 240 timed trials per subject.

2.1.2 Dependent Variables

The primary dependent measure was response time (RT), measured from the time the target was presented until a target identifier was typed. Subjects were allowed to study each screen layout for as long as they wished before starting the experimental trials; the viewing time for each condition was recorded but not analyzed because subjects were allowed to ask questions during this period. After the experimental trials, the rank order of preference for combinations of Color and Texture coding for three criteria were collected: subjective discriminability, visual appeal, and overall preference.

2.2 Apparatus and Stimuli

Stimuli were presented on a dedicated Hewlett-Packard color workstation, and responses were recorded with two keys on its keyboard. Coded windows were displayed on a rectangular medium gray background called the presentation area (see Figure 1). Each window had a two-part name: an application name and a visually distinct single-letter identifier within an application (X, O, W, B). There were 4 Edit windows (Edit_X, Edit_O, Edit_W, Edit_B), 3 News windows (News_X, News_O, News_W), 2 Mail windows (Mail_X, Mail_O), and 1 Draw window (Draw_X).

On each trial, the name of the target window was presented below the presentation area (Mail_O in Figure 1). The name was surrounded by the coding of the target window (20-pixel dots in Figure 1); we wanted to test the discriminability of different codes, not subjects' abilities to recall the codes because if the codes were not discriminable, their memorability would not matter.

To ensure that subjects were accurately finding the targets, randomly chosen left or right angle brackets were placed around the names on each window name (right bracket in Figure 1); when subjects found the target, they pressed the left or right arrow key assigned to the target window on that trial. This procedure had been found to be effective by Perlman (1984).

2.3 Subjects

Sixteen paid subjects (14 male, 2 female), with normal-corrected vision without color-blindness, were drawn from a pool of university students.

2.4 Procedure

Subjects were told to respond as quickly and accurately as possible. Subjects were given 10 practice trials to become familiar with the scheme of finding a target and typing the target identifier (a left or a right angle-bracket). Subjects did not have problems learning the technique. The timed trials took about 30 minutes.

At the end of the timed trials, subjects were given a questionnaire in which they ranked their preferences for the different coding methods and indicated pros and cons of the different methods. Subjects were explained the general purpose of the experiment and offered a copy of the final report on the data.

Originally published: Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, 1993, pp. 343–347. **Republished**: G. Perlman, G. K. Green, & M. S. Wogalter (Eds.) *Human Factors Perspectives on Human-Computer Interaction: Selections from Proceedings of Human Factors and Ergonomics Society Annual Meetings*, 1983–1994, Santa Monica, CA: HFES, 1995, pp. 321-325.

3 RESULTS

The results show that color is an effective coding method, especially when same-color-coded windows are widely dispersed. Because of the high power of the experimental design (240 points per subject in a largely within-subjects design), some effects were significant, but considered unimportant because of the small size of the effects. In those cases, the index of association, ω^2 (Keppel, 1973), is reported to indicate the percentage of variance accounted for by the effects observed. As a general rule, we considered any effect accounting for less than 1% of the variation as unimportant.

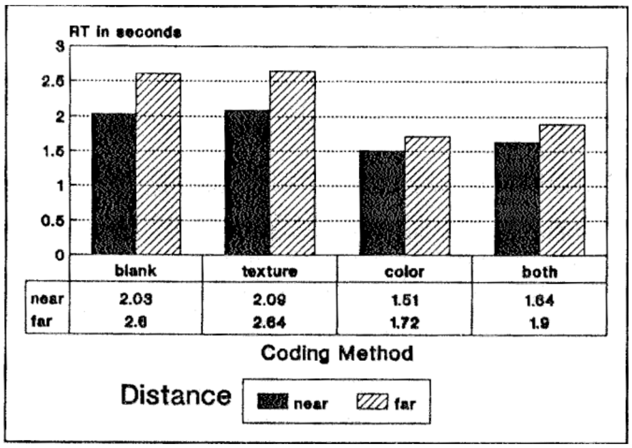

Figure 2. Response times for coding methods for near and far window configurations. Average standard error is 42 msec.

Response Time: The main results are presented in Figure 2. Color coding significantly reduced response time (RT) ($F(1,15)=138.63$, $p<.001$), but Texture had no effect ($F(1,15)=2.78$, $p>.10$). Although it appears that Color + Texture coding was worse than color coding alone, the interaction of Color and Texture was not significant ($F(1,15)=1.49$, $p>.2$). Distance between windows significantly increased RT ($F(1,15)=45.29$, $p<.001$), and there was a significant Color by Distance interaction ($F(1,15)=24.06$, $p<.00l$); search for Color coded targets was less affected by increased Distance.

Figure 3 shows the effect of Group size on RT. As the number of targets increased, so did RT ($F(3,45)=21.04$, $p<.001$). There was a small but significant Distance x Group size interaction ($F(3,45)=6.14$, $p<.01$), ω^2 was .006 and contributed to one eighth of the variance as the Group size factor. There was no effect of Trial ($F(9,135)=1.59$, $p>.1$), although the first trial in a block of 10 was over 100 msec longer than any of the others. There was no indication (from the mean or the variance) of a drop for the last trial in a block; such a drop would have indicated that subjects were able to recall the previous nine trials and predict the final trial.

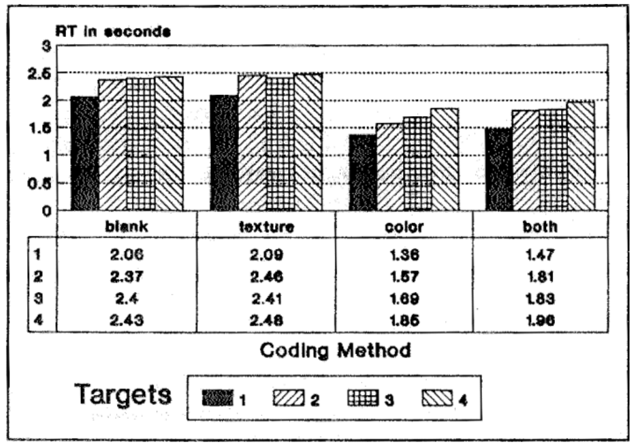

Figure 3. Response times for coding methods for window-group sizes. Average standard error is 61 msec.

There was no significant effect of Border width ($F(2,30)<1$) and the range of times for the three Borders widths was less than 36 msec, with a 35 msec average standard error. There was a small but significant Color x Border width x Group size interaction ($F(6,90)=2.58$, $p<.05$), ω^2 was .002. There was a small but significant effect of Replication ($F(2,30)=10.87$, $p<.001$), ω^2 was .004; the first replication was 87 msec slower than the second which was 20 msec slower than the third.

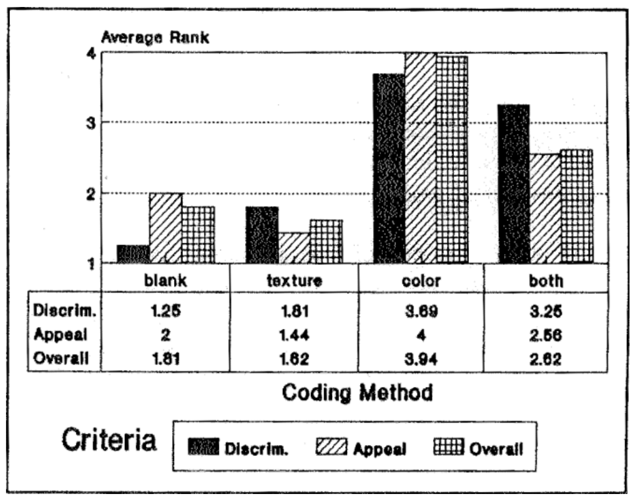

Figure 4. Subjective ratings of coding methods. Higher rankings indicate higher subjective discriminability, visual appeal, and overall preference.

Ratings: Subjects' rank orderings of coding methods (low is bad, high is preferred) for three criteria: discriminability, visual appeal, and overall preference were analyzed as rank-ordinal data. The average ranks are shown in Figure 4. The results mirror the response time data. A Friedman chi-square test of average ranks showed significant variation for each question at the .001 level.

Originally published: Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, 1993, pp. 343–347. **Republished**: G. Perlman, G. K. Green, & M. S. Wogalter (Eds.) *Human Factors Perspectives on Human-Computer Interaction: Selections from Proceedings of Human Factors and Ergonomics Society Annual Meetings*, 1983–1994, Santa Monica, CA: HFES, 1995, pp. 321-325.

Errors: If subjects were guessing the left/right code for the windows, the expected error rate would be 50%. There were 77 errors in 3840 trials, resulting in an error rate of 2%. This indicates that the subjects were finding the targets before responding. No analysis of number of errors showed significant effects, and only the number of errors by Group size was interesting (4, 14, 26, and 33 for Group sizes 1, 2, 3 and 4, respectively). The mean time for errors was 1.91 seconds, compared to 2.01 seconds overall, suggesting a small speed-accuracy tradeoff. An analysis of the 49 response times longer that 5.0 seconds (an arbitrarily chosen cutoff) showed that 47/49 were in the no-color condition and 42/49 were in the far configuration (40/49 were in the no-color, far configuration).

4 DISCUSSION

The effectiveness of color coding has already been demonstrated, so these results will serve mainly to provide designers with an empirically valid performance-based reason to avoid texture coding. Not only can it be annoying, as Tufte (1983) points out, but it does not appear to be an effective method to make groups of similar objects stand out. For distant targets, the advantage of color coding over no coding was 873 msec, while the difference of texture and no coding was negligible (and in the wrong direction).

Statistical power was not a problem in this experiment. Given that the average standard error for the coding methods in Figure 2 was 42 msec, and that the experiment was powerful enough to detect differences of about 100 msec (which is far less than the effect of color coding), it is safe to conclude that the effectiveness of texture coding is at best drastically less than color coding. A Scheffé 99% confidence interval around the (dis)advantage of Texture coding was 100±177 msec. The ratio of the ω^2 values for Color vs. Texture was about 50-to-1.

We do not believe that a different set of textures would be more effective because pilot studies had shown no differences among different sets, and the codes used in this experiment were about as different as possible without making it an experiment studying the effectiveness of brightness coding. (We believe that different densities would be a more effective coding method, but we have not yet collected that data.) Based on comments by subjects that the dots pattern (see Figure 1) was easy to find, we analyzed the data to see if there was significant variation among the textures. The four textures were within 74 msec of each other, and did not differ significantly; the four colors were within 110 msec of each other and also did not differ significantly.

We plan to follow up this experiment with one studying the effectiveness of different textures, densities, and possibly spatial frequencies in graphs like those in Figures 2–4. Some reviewers of this paper have questioned whether the textures interfered with the labels for which subjects were searching, so a task that required subjects to look for trends and not labels (e.g., of Group-size 3 in Figure 3) could take advantage of an effective coding method but avoid this possible confounding effect.

5 REFERENCES

Card, S. K and Henderson, A, Jr. (1987) A Multiple, Virtual-Workspace Interface to Support User Task Switching. **Proceedings of ACM CHI+G1'87 Conference on Human Factors in Computing Systems and Graphics Interface**, 53–59.

Galitz, W. O. (1993) **User-Interface Screen Design**. Wellesley, MA: Q. E. D. Information Sciences.

Helander, M. G. (1987) Design of Visual Displays. In G. Salvendy (Ed.) **Handbook of Human Factors**, 507–548. New York: John Wiley & Sons.

Keppel, G. (1973) **Design and Analysis: A Researcher's Handbook**. Englewood-Cliffs, NJ: Prentice-Hall.

Perlman, G. (1984) Making the Right Choices with Menus. **Proceedings of IFIP INTERACT '84: Human-Computer Interaction**, 317–321.

Rice, J. F. (1991) Display Color Coding: 10 Rules of Thumb. **IEEE Software**, 8:1, 86–88.

Smith, S. L and Mosier, J. N. (1986) **Guidelines for Designing User Interface Software** (ESD-TR-86-278). Bedford, MA: The MITRE Corporation.

Swierenga, S. J., Boff, K R. and Donovan, R. S. (1991) Effectiveness of Coding Schemes in Rapid Communication Displays. **Proceedings of the Human Factors Society 35th Annual Meeting**, 1522–1526.

Tufte, E. R. (1983) **The Visual Display of Quantitative Information**. Cheshire, CT: Graphics Press.

Ware, C. and Knight, W. (1992) Orderable Dimensions of Visual Texture Useful for Data Display: Orientation, Size, and Contrast. **Proceedings of ACM CHI '92 Conference on Human Factors in Computing Systems**, 203–209.