

Human Factors: The Journal of the Human Factors and Ergonomics Society

<http://hfs.sagepub.com>

A Model of Clutter for Complex, Multivariate Geospatial Displays

Maura C. Lohrenz, J. Gregory Trafton, Melissa R. Beck and Marlin L. Gendron

Hum Factors 2009; 51; 90

DOI: 10.1177/0018720809333518

The online version of this article can be found at:
<http://hfs.sagepub.com/cgi/content/abstract/51/1/90>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Human Factors and Ergonomics Society](http://www.humanfactors.org)

Additional services and information for *Human Factors: The Journal of the Human Factors and Ergonomics Society* can be found at:

Email Alerts: <http://hfs.sagepub.com/cgi/alerts>

Subscriptions: <http://hfs.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://hfs.sagepub.com/cgi/content/refs/51/1/90>

A Model of Clutter for Complex, Multivariate Geospatial Displays

Maura C. Lohrenz, Naval Research Laboratory, Stennis Space Center, Mississippi, J. Gregory Trafton, Naval Research Laboratory, Washington, D.C., Melissa R. Beck, Louisiana State University, Baton Rouge, Louisiana, and Marlin L. Gendron, Naval Research Laboratory, Stennis Space Center, Mississippi

Objective: A novel model of measuring clutter in complex geospatial displays was compared with human ratings of subjective clutter as a measure of convergent validity. The new model is called the *color-clustering clutter* (C3) model. **Background:** Clutter is a known problem in displays of complex data and has been shown to affect target search performance. Previous clutter models are discussed and compared with the C3 model. **Method:** Two experiments were performed. In Experiment 1, participants performed subjective clutter ratings on six classes of information visualizations. Empirical results were used to set two free parameters in the model. In Experiment 2, participants performed subjective clutter ratings on aeronautical charts. Both experiments compared and correlated empirical data to model predictions. **Results:** The first experiment resulted in a .76 correlation between ratings and C3. The second experiment resulted in a .86 correlation, significantly better than results from a model developed by Rosenholtz et al. Outliers to our correlation suggest further improvements to C3. **Conclusions:** We suggest that (a) the C3 model is a good predictor of subjective impressions of clutter in geospatial displays, (b) geospatial clutter is a function of color density and saliency (primary C3 components), and (c) pattern analysis techniques could further improve C3. **Application:** The C3 model could be used to improve the design of electronic geospatial displays by suggesting when a display will be too cluttered for its intended audience.

INTRODUCTION

Clutter is a known problem in electronic geospatial displays, on which many types of data (multivariate) are fused and presented as one image. Displays ranging from handheld global positioning systems to complex geospatial information systems (GIS) and military moving maps present vast amounts of information, including maps, weather radar, satellite imagery, routes, place names, and metadata (descriptive information such as runway lengths and urban populations). In this context, *clutter* can refer to unwanted or unnecessary information, but often it refers to an overabundance of useful information.

Excess clutter on a display can result in confusion by obscuring important information. Previous studies have associated user performance with display complexity; for example, clutter on an aircraft moving map has been found to disrupt a pilot's visual attention, resulting in greater uncertainty concerning target locations (Schons & Wickens, 1993; Wickens & Carswell, 1995). Lohrenz, Myrick, Trenchard, Ruffner, and Cohan (2000) reported that pilots will turn off the moving-map display if they perceive it to be too cluttered, thereby removing a potential distraction.

Research has identified several factors related to clutter that increase search time to

Address correspondence to Maura C. Lohrenz, Naval Research Laboratory, Code 7440.1, Stennis Space Center, MS 39529; mlohrenz@nrlssc.navy.mil. *HUMAN FACTORS*, Vol. 51, No. 1, February 2009, pp. 90-101. DOI: 10.1177/0018720809333518. Copyright © 2009, Human Factors and Ergonomics Society.

find a predetermined target. For example, a target is harder to locate as the number of items in a visual array that share features with the target increases (Wolfe, 1998). The similarity of the target to other information in the visual array also affects search performance, with higher degrees of similarity leading to slower search (Duncan & Humphreys, 1989). Search becomes less efficient as items are positioned more closely together, allowing them to occlude one another (Bravo & Farid, 2004a, 2004b). When items are arranged in a predictable fashion (e.g., a grid), targets are found more readily than when items are randomly arranged (Beck & Trafton, 2007). Finally, targets are harder to locate in more complex backgrounds (Neider & Zelinsky, 2006). Considering all the factors that can affect the relationship between clutter and visual search, the usefulness of a quantifiable clutter model becomes readily apparent.

The obvious first step in managing display clutter is determining whether a display is, in fact, too cluttered, which requires a reliable definition of clutter. Clutter has been defined extensively for text-based displays. Tullis (1981, 1983) found that both people's perception of clutter and their ability to locate items in screens of text are influenced by (a) the density of text on the screen, (b) the number and size of text groups, (c) the complexity of text groups, and (d) the presence of highlighting using color, size, and so on.

Tullis's (1981, 1983) definition can be extended to multivariate displays consisting of geospatial features instead of (or in addition to) text. In GIS terms, a *feature* is any displayed item categorized as one of four types: a point with an associated symbol (e.g., airport location or city center), a line (road or river), an area (forest or urban area), or text (city name or elevation label). Following Tullis, we suggest that clutter in a geospatial display is influenced by (a) the density of discernible features on the display, (b) the number and size of features, (c) the complexity or diversity of features, and (e) feature highlighting using color, size, and so on.

We are interested in the impact of clutter on the effectiveness of a display, measured by a user's assessment of clutter in the display and his

or her ability to access and use information. For this reason, the definition of clutter proposed by Rosenholtz, Mansfield, and Jin (2005) is appropriate for our purposes: Clutter is the state in which excess items, or their representation or organization, lead to a degradation of performance at some task. Rosenholtz et al. suggested that clutter is a function of how easily a new, salient item could be added to a display. According to their paradigm, cluttered displays have an overly congested feature space, leaving little room for new features. Their model measures clutter as a function of *feature congestion*, whereby the features in question are color and luminance contrast rather than actual displayed items.

Other researchers (e.g., Zuschlag, 2004) who have attempted to quantify clutter in complex displays also focus on the contribution of luminance contrast or saliency to clutter. *Saliency*, in this case, refers to how conspicuous a displayed item appears to the user relative to other items in the display (modeled as a *saliency map* in Itti, Koch, & Niebur, 1998). In the context of luminance contrast, saliency is the difference in luminance between the item of interest and surrounding items or background. Saliency can be thought of as a form of highlighting in Tullis's (1981, 1983) list of contributions to clutter. We recognize different scales of saliency, from the saliency of one displayed item ("local" saliency) to average saliency over an entire image ("global" saliency).

OUR MODEL OF VISUAL CLUTTER

We suggest that our perception of clutter in complex geospatial displays is influenced by an interaction between global saliency and color density, which can be thought of as color homogeneity—the opposite of color variability. We define *color density* as a measure of how tightly packed similar-colored pixels are within an image. At one extreme, an image consisting of one solid color has the maximum color density of 1.0 (100%). Adding scattered pixels of different colors would lower color density.

Color density relates to two additional clutter components in Tullis's (1981, 1983) list: feature density and display complexity. In general, we theorize that low saliency plus high color density

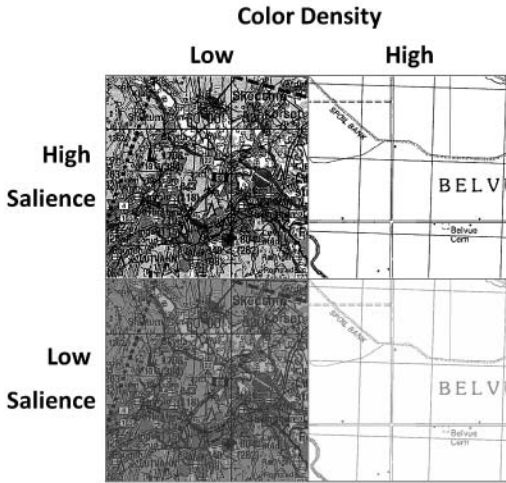


Figure 1. Electronic chart samples illustrating the impacts of color density and global saliency on clutter.

results in the impression of low clutter, whereas high saliency plus low color density results in the impression of high clutter. However, although increasing saliency in an image with low color density increases the impression of clutter (Figure 1, left two images), increasing saliency in an image with very high color density has little or no effect on clutter (Figure 1, right two images).

We propose the color-clustering clutter (C3) model, which uses a series of algorithms to detect visually distinct features in an image as clusters of similar-colored pixels, compute each cluster's color density, compute saliency between clusters, and calculate clutter as a function of color density and saliency.

Color Density

We compute color density by clustering all of an image's pixels in both geospatial location and color difference, such that adjacent pixels with similar colors cluster together, and calculating the density of pixels per cluster. Each cluster represents a visually discernible feature in the display.

The color-clustering algorithm was adapted from a point-clustering algorithm (Gendron, Layne, & Lohrenz, 2005), in which points to be clustered are represented in a so-called geospatial bitmap. Each point is represented by a "set" bit (value = 1), and all other bits in the bitmap

are "cleared" (0). The algorithm expands each set bit into a predetermined shape. All expanded bits that touch or overlap are clustered via the formation of new bitmaps. After clustering, another algorithm (Layne, Gendron, & Lohrenz, 2008) produces vertices for a bounding cluster polygon, the density of which is the number of clustered points divided by the polygon's area.

During color clustering, image pixels are clustered if they are close (within predefined thresholds) with respect to both geospatial location (x, y) and color difference (z). The x and y thresholds are each 1 pixel, and the z threshold is chosen to approximate a just noticeable difference between colors, according to an appropriate color difference formula (discussed later).

The algorithm operates on each color in the image, starting with the most prevalent color. The density of each color being clustered (the "seed" color) is the weighted average (by area) of the densities of all cluster polygons containing that color, where the density D_p of a cluster polygon p is as follows:

$$D_p = \sum_{c=1 \text{ to } n} [(1 - E_c/M)N_c]/A_p,$$

where n = number of colors in the image within distance z of seed color, E_c = Euclidean distance between color c and seed color in chosen color space, M = maximum possible color difference in chosen color space, N_c = number of pixels of color c in polygon p , and A_p = area of p .

Each pixel is included in a single cluster: After inclusion in a cluster, the pixel is removed from the list of pixels yet to be clustered. If a pixel lies inside a cluster polygon's boundary, but the difference between its color and that cluster's seed color is greater than z , then the pixel is excluded from that cluster's density calculation and will be clustered with another seed color.

The color density D of the image is a weighted average (by total area) of all the polygon densities D_p in the image. Lohrenz and Gendron (in press) describe this algorithm in more detail.

Saliency

After clustering has been completed, saliency S is computed as a weighted average (by edge length) of color differences among adjacent clusters:

$$S = \Sigma(Z_e L_e) / \Sigma(L_e),$$

where Z_e = difference in color between adjacent clusters along their shared edge e and L_e = length of e in linear pixels.

Color Difference Formula

For both color density and saliency, z depends on the color difference formula used. The Commission Internationale de l'Eclairage (CIE) developed the Δ Lab color difference formula (Euclidean distance between colors in CIE $L^*a^*b^*$ space) and the more recent CIEde2000 (CIE, 2001) formula to approximate the human visual system's ability to detect color differences. However, neither agrees perfectly with psychophysical color difference data, and CIEde2000 in particular is very complex and computationally intensive (Sharma, 2003).

Other color difference formulas we considered include Δ HSV (Euclidean distance in hue, saturation, and value), Δ RGB (red, green, blue), and a preliminary " Δ RGBL" formula that we propose in this article, in which the tristimulus RGB values are weighted by the human eye's relative sensitivities to red, green, and blue light, estimated by the luminosity response curve in Foley and Van Dam (1984, Figure 17.18), scaled to luminance (L) = 1:

$$\Delta\text{RBL} = \text{SQRT}[(\Delta R * 0.3)^2 + (\Delta G * 0.6)^2 + (\Delta B * 0.1)^2 + \Delta L^2] * \text{MAX}(L1, L2) / 100.$$

We performed an experiment to choose which formula was best for calculating z in the color-clustering algorithm (i.e., to determine whether adjacent pixel colors are just noticeably different). Participants were shown a series of solid-colored squares in which a small dot, close in color to the background, was randomly placed within a 2° visual angle from the center of the display (10% of stimuli contained no dot). Dot colors were chosen to be similar to the background in one color component (R, G, B, L, a, b, H, S, or V). Participants were asked to report the following: 1 = *cannot see a dot*, 2 = *can barely see a dot*, or 3 = *can clearly see a dot*.

A series of linear regression analyses suggest that the color difference formula that best

predicted participants' ability to see the dot was our proposed Δ RGBL, with a correlation (r) of 0.55 ($r^2 = .30$; $t = 22.2$, $n = 1,125$, $p < .0001$). The other formulas had less impact on participants' ability to see the dot, although all t tests were significant ($p < .0001$): Δ RGB $r = .39$, Δ Lab $r = .36$, CIEde2000 $r = .34$, and Δ HSV $r = .17$. Results were comparable on each of four identical computer monitors set to factory-default brightness settings.

The color-clustering algorithm clusters adjacent pixels if they are close in color. Because the algorithm considers each pixel individually, this process is similar to determining color closeness for the small dots in the previous experiment. The best color difference formula in this case was Δ RGBL. For Δ RGBL values greater than 12, participants could consistently (but barely) see a difference (mean response = 2.1). Therefore, we chose $z = 12$ Δ RGBL units as an estimate of just-noticeable color differences for the clustering algorithm.

C3

After clustering all pixels in an image and computing color density (D) and global saliency (S), averaged across all clusters, we found that the C3 model estimates clutter as a continuous series of exponential growth curves (Figure 2):

$$\text{Clutter} = F(1-D)e^{-Ie^{-S/G}}$$

F is an arbitrary scale factor, and $F(1 - D)$ is the curve's upper asymptote (i.e., maximum C3 value) for color density D . We set $F = 15$ to map the clutter values to the subjective clutter scale used in the following experiments. The parameters I and G , which determine the curve's inflection point and (with saliency) growth rate, respectively, were derived empirically in one experiment and validated in a second experiment.

Our clutter model is an adaptation of the Gompertz function (Gompertz, 1825), which is used to model time series data in which growth is slowest at the beginning and end of a period. Waliszewski and Konarski (2005) explained that this function models "the co-existence of at least two antagonistic processes with the complex coupling of their probabilities" (p. 277).

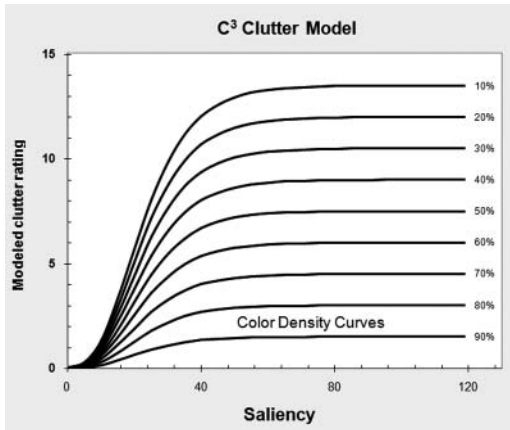


Figure 2. A continuous series of curves adapted from the Gompertz growth function defines the color-clustering clutter model.

In our case, the two processes are color density and saliency.

Figure 2 illustrates how this function fits our concept of clutter: For very high color densities (e.g., 90%), clutter remains low regardless of saliency (as previously shown in Figure 1, right). As color density decreases, saliency has a potentially greater impact on clutter. Meanwhile, with low global saliency (below about 50 in this model), adjacent features tend to be more difficult to distinguish, reducing the amount of perceived clutter. At higher saliencies, adjacent features become clearly discernible, and clutter becomes a linear function of color density.

As an initial test of the convergent validity of the C3 model, the following experiments focus on subjective impressions of clutter, which we compare with both C3 and Rosenholtz clutter measures.

CLUTTER EXPERIMENT 1

In Experiment 1, we manipulated the two C3 variables (I and G) to maximize correlations between our model and subjective impressions of clutter for a wide variety of graphic displays. We verified these variables using new stimuli in Experiment 2.

Method

Participants. From among those we recruited, 57 undergraduate psychology students from

George Mason University participated in this study for course credit. Participants ranged in age from 18 to 26 ($M = 20.1$) years. All had normal or corrected-to-normal vision; none was color-blind. The experiment lasted approximately 30 min.

Materials. Stimuli consisted of 58 diverse graphic displays in six categories: airport terminal maps, flowcharts, road maps, subway maps, topographic charts, and weather maps (Figure 3). Most graphics contained multiple feature types (points, lines, areas, text) and colors. Graphics ranged in size from 300×600 pixels to 800×600 pixels and were displayed at 72 pixels per inch. Five graphics were downloaded from Rosenholtz's Web site (<http://web.mit.edu/ruth/>) and are described in Rosenholtz et al. (2005).

Procedure. Participants were tested in small groups of 5 to 10. Each group received a different random order of stimuli. Participants were asked to rate how cluttered they thought each graphic appeared, from 1 (*not cluttered*) to 10 (*extremely cluttered*). Participants were given as much time as needed to rate each graphic; most responded within 10 s.

Results

We calculated global saliency and color density for each graphic, then manipulated the two C3 variables and compared resultant clutter values with subjective ratings. The highest correlations (within graphic categories and for all graphics combined) occurred when $I = 6.3$ and $G = 10$.

Average clutter ratings of stimuli ranged from 3.3 to 9.5 ($M = 6.2$, $SD = 1.5$). Overall, participants were relatively consistent with one another, $r = .47$, $z = 3.5$, $p < .0005$. To explore how well C3 matched subjective ratings, the mean rating of each graphic was calculated and correlated with C3, shown in Table 1 (averaged by category) and Figure 4. The average C3 for stimuli ranged from 1.2 to 8.7 ($M = 3.4$, $SD = 1.8$).

We also calculated clutter for each graphic using software featured in Rosenholtz et al. (2005), available from <http://web.mit.edu/ruth/www/#software>, and correlated these values with mean ratings (also shown in Table 1 and

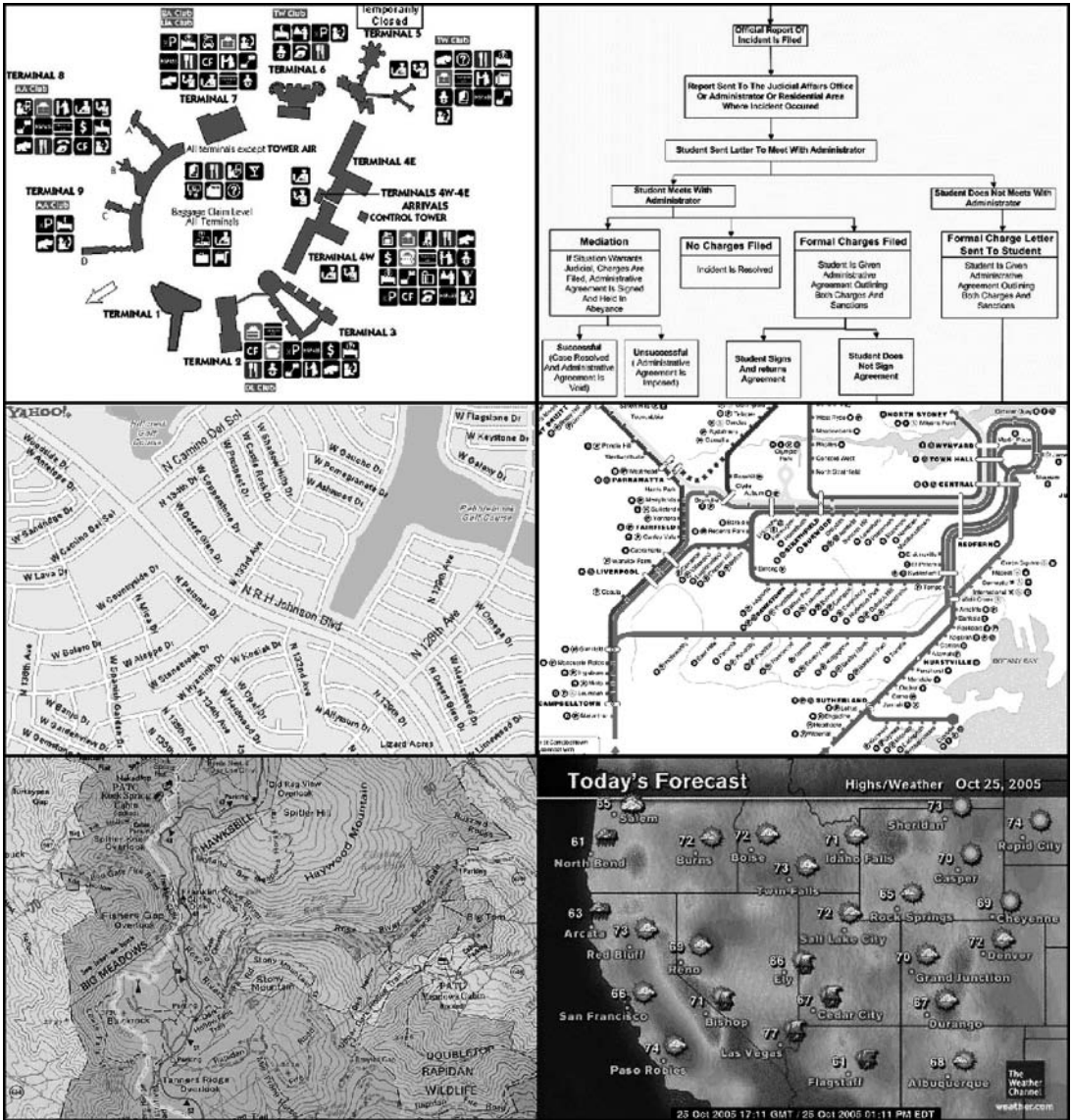


Figure 3. Examples of graphic displays (by category) shown to participants during Experiment 1. Top: airport terminal maps, flowcharts; middle: road maps, subway maps; bottom: topographic charts, weather maps.

Figure 4). Average Rosenholtz clutter values for stimuli ranged from 0.2 to 0.6 ($M = 0.3$, $SD = 0.1$).

Not surprisingly, different types of graphics varied in their mean subjective ratings. Participants rated airport terminal maps, flowcharts, and weather maps approximately 1 point less cluttered than road maps, subway maps, and topographic charts. Both C3 and Rosenholtz's

models also found airport terminal maps, flowcharts, and weather maps to be less cluttered, on average, than the other three categories.

Discussion

The C3 model accounts for 58% of the variance ($r = .76$, $t = 8.8$, $p < .0001$) in subjective clutter ratings for all graphics combined, whereas Rosenholtz's model accounts for 47%

TABLE 1. Correlations Between Mean Ratings and Clutter Models, Averaged by Graphic Category

Graphic Category (Number of Graphics)	Mean Rating	Mean C3	C3 Correlation (<i>r</i>) With Ratings	Mean Rosenholtz Clutter	Rosenholtz Correlation (<i>r</i>) With Ratings
Airport terminals (9)	5.87	2.10	.48 ($p > .10$)	0.30	.11 ($p > .10$)
Flowcharts (8)	5.70	2.71	.78 ($p = .02$)	0.28	.72 ($p = .04$)
Road maps (14)	6.87	4.39	.82 ($p = .0003$)	0.33	.90 ($p < .0001$)
Subway maps (9)	6.70	3.13	.81 ($p = .008$)	0.33	.54 ($p > .10$)
Topographical charts (8)	6.81	5.37	.81 ($p = .014$)	0.37	.68 ($p > .05$)
Weather maps (10)	5.20	2.26	.81 ($p = .004$)	0.22	.55 ($p > .05$)
All graphics (58)	6.23	3.38	.76 ($p < .0001$)	0.31	.68 ($p < .0001$)

Note: C3 = color-clustering clutter.

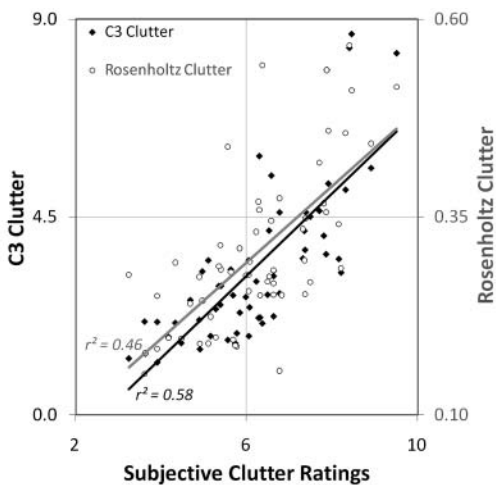


Figure 4. The correlation between mean subjective ratings of clutter and color-clustering clutter metric for all graphics in Experiment 1 is .76. The correlation between ratings and Rosenholtz's clutter metric for these graphics is .68.

of the variance ($r = .68$, $t = 7.0$, $p < .0001$). Fisher's r to z' transformation did not show these two correlations to be significantly different ($z = 1.3$, $p = .90$), which suggests that both models did a similarly good job predicting subjective impressions of clutter in this data set. It is remarkable that both models accounted for approximately 50% of variability in terms of subjective clutter.

The C3 model also correlates very well ($r \geq 0.78$, $p < 0.03$) with mean ratings for all but one graphic category (airport terminal maps),

suggesting that C3 is an appropriate model of clutter for many different graphic types. This is partly because we manipulated our free variables to maximize correlations both within categories and for all graphics combined, in hopes that C3 would be robust for many diverse graphic types. In contrast, correlations between mean ratings and the Rosenholtz model are considerably lower than C3 for all but one category (road maps), and only two categories (road maps and flowcharts) resulted in $p < .05$, which suggests that this clutter model might be appropriate for a more limited set of graphic types.

Most graphics in this experiment consisted of a single, dominant feature type, and the number of these features (the feature set size) seems to have significantly influenced subjective clutter ratings. For example, the flowcharts consisted entirely of text-based "bubbles," the number of which could be thought of as each flowchart's feature set size. Likewise, each airport terminal map consisted almost entirely of similar-looking icons, each weather map mostly consisted of pairs of weather icons and temperature labels, and the road and subway maps consisted of linear features (roads, subway lines) and associated text strings. The correlation between feature set size and subjective ratings for these displays was .58 ($r^2 = .33$), suggesting that up to a third of the variation in ratings could be explained by the number of features in each image.

Despite a significant set size effect, Experiment 1 showed strong support for both clutter



Figure 5. Samples of aeronautical chart images shown to participants during Experiment 2. Top: low clutter bin; middle: medium clutter bin; bottom: high clutter bin.

models, with C3 slightly stronger and potentially a more versatile model of clutter than Rosenholtz's for diverse graphic categories. The C3 model, however, used two free parameters determined empirically from this data set. To determine whether those parameters would generalize to other graphics, we ran a second experiment with different participants and different graphics while keeping the two C3 parameters constant.

EXPERIMENT 2

Experiment 2 is a replication of Experiment 1 and a validation of the C3 parameters defined

in Experiment 1. Experiment 2 also was designed to minimize the set size effect seen in Experiment 1 by using graphics with more diverse, dissimilar features.

Method

Participants. From among those we recruited, 55 undergraduate psychology students from Louisiana State University participated in this study for course credit. Participants ranged in age from 18 to 33 ($M = 20.1$) years. All participants had normal or corrected-to-normal vision; none was color-blind. The experiment lasted approximately 10 min.

TABLE 2. Mean Subjective Ratings and Clutter Models for Each Graphic (and Averaged for Each Clutter Category) in Experiment 2

Chart	M (SD) Rating	C3	Rosenholtz Clutter
L01	2.1 (1.3)	1.5	0.23
L02	1.7 (1.6)	1.5	0.21
L03	1.9 (1.3)	1.9	0.20
L04	1.5 (1.4)	2.0	0.23
L05	2.2 (1.6)	2.0	0.20
L06	1.6 (1.4)	2.0	0.19
L07	2.0 (1.3)	2.1	0.32
L08	2.0 (1.6)	2.1	0.30
L09	1.4 (1.2)	2.1	0.24
L10	2.9 (2.1)	2.4	0.39
L11	2.0 (1.6)	2.4	0.21
L12	0.8 (0.9)	3.0	0.32
L13	2.7 (1.8)	3.2	0.28
L14	2.0 (1.6)	3.4	0.25
L15	2.9 (1.8)	3.5	0.32
L16	3.2 (1.6)	3.5	0.30
L17	2.2 (1.4)	3.7	0.26
L18	3.3 (1.9)	4.1	0.35
M01	7.0 (1.7)	5.6	0.3
M02	6.9 (1.7)	5.7	0.4
M03	4.1 (1.9)	5.8	0.3
M04	6.0 (1.7)	5.8	0.4
M05	5.5 (1.8)	5.8	0.3
M06	4.8 (2.1)	5.9	0.4
M07	8.5 (1.1)	5.9	0.4
M08	7.2 (1.7)	6.0	0.6
M09	5.1 (1.8)	6.0	0.4
M10	6.6 (2.3)	6.0	0.4
M11	6.5 (1.8)	6.1	0.4
M12	5.7 (1.6)	6.1	0.4
M13	5.3 (1.6)	6.2	0.3
M14	6.7 (1.5)	6.2	0.4
M15	6.5 (1.5)	6.2	0.3
M16	7.4 (1.5)	6.3	0.5
M17	6.1 (1.4)	6.3	0.3
M18	6.9 (1.9)	6.3	0.4
Z01	6.8 (1.6)	7.8	0.4
Z02	8.1 (1.4)	7.9	0.5
Z03	7.3 (1.4)	8.0	0.4
Z04	6.2 (1.9)	8.1	0.4
Z05	8.2 (0.9)	8.1	0.5
Z06	8.0 (1.2)	8.2	0.5

(continued)

TABLE 2. (continued)

Chart	M (SD) Rating	C3	Rosenholtz Clutter
Z07	7.8 (1.4)	8.2	0.5
Z08	8.2 (1.0)	8.3	0.4
Z09	7.6 (1.5)	8.4	0.4
Z10	5.6 (1.7)	8.4	0.5
Z11	7.7 (1.3)	8.4	0.4
Z12	7.9 (1.3)	8.5	0.4
Z13	6.7 (1.8)	8.7	0.5
Z14	5.7 (1.7)	8.8	0.5
Z15	7.7 (1.2)	8.8	0.4
Z16	6.2 (2.0)	8.9	0.5
Z17	3.2 (2.0)	9.0	0.4
Z18	8.4 (1.1)	9.3	0.6

Note: C3 = color-clustering clutter.

Materials. Stimuli consisted of 54 aeronautical chart samples presented as 256-color GIF images (740 × 580 pixels), including 21 civilian sectionals and 33 military aeronautical charts from the National Geospatial-Intelligence Agency. Each graphic subtended 28.5° × 23.2° (computed from a viewing distance of 35 cm).

We initially reviewed 175 chart samples before choosing the final set, in an attempt to include as wide a range of clutter as possible while maintaining a diversity of chart types and color palettes. We ran our C3 model (using the parameters set in Experiment 1), recorded a clutter value for each of the 175 charts, and chose 18 charts (of varying chart series) to represent each of three clutter bins: low, medium, and high (Figure 5).

We also ran the Rosenholtz clutter model on the 54 charts to ensure that those clutter values had a similar distribution. Not all charts fell into the same clutter bins for both models. However, if we assume for each model that the lowest 18 clutter values are “low,” the middle 18 are “medium,” and the highest 18 are “high,” then normalized (from 0 to 1) clutter averages (and standard deviations) are fairly consistent between the two models:

C3: low, 0.14 (SD 0.10); medium, 0.58 (0.03); high, 0.89 (0.05); overall, 0.54 (0.32).

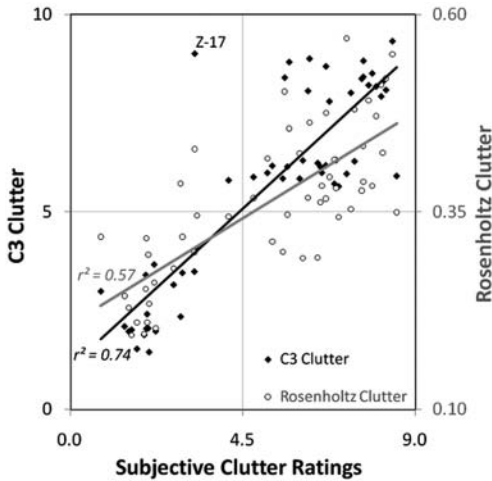


Figure 6. The correlation between mean subjective ratings of clutter and color-clustering clutter (C3) metric for all graphics in Experiment 2 is .86. The correlation between ratings and Rosenholtz's clutter metric for these graphics is .75. The C3 value for Chart Z17 is an outlier.

Rosenholtz: low, 0.17 (SD 0.11); medium, 0.44 (0.06); high, 0.73 (0.13); overall, 0.45 (0.25).

Procedure. Participants were tested in groups of 4 or fewer. Using SuperLab 4.0.2 (Cedrus Corporation), we presented stimuli to each participant in random order and recorded responses. Presentation of graphics was self-paced for each participant. As soon as a response was given, the next graphic appeared on the participant's display. Participants first completed a target search task on each graphic, presented in random order, results of which are published in Beck, Lohrenz, Traflet, and Gendron (2008). Following this task (which took an average of 17 min per participant), participants took a 5-min break (on average) and then were shown the same graphics in a different random order and asked to rate how cluttered they thought each appeared, from 0 (*not cluttered*) to 9 (*extremely cluttered*). Most participants answered in less than 3 s per graphic.

Results

Average clutter ratings of stimuli ranged from 0.8 to 8.5 ($M = 5.2$, $SD = 2.4$). The mean rating for each graphic was compared with

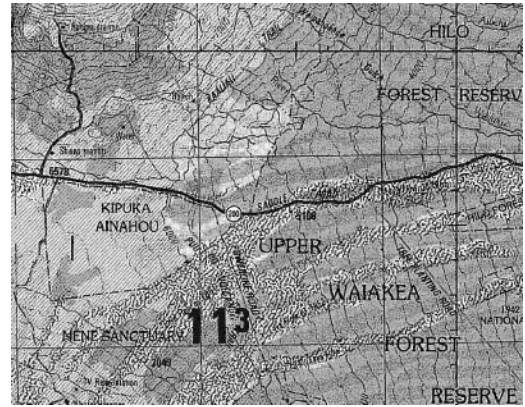


Figure 7. Chart resulting in outlier color-clustering clutter (C3) rating. Participants rated Chart Z17 relatively uncluttered (3.2), but the C3 model rated it extremely cluttered (9.0).

C3 and Rosenholtz clutter values (Table 2 and Figure 6): Average C3 values ranged from 1.5 to 9.3 ($M = 5.7$, $SD = 2.5$), and average Rosenholtz clutter values ranged from 0.2 to 0.6 ($M = 0.4$, $SD = 0.1$).

Both clutter models correlate very well with mean ratings for these graphics: C3 $r = .86$ ($p < .0001$) and Rosenholtz $r = .75$ ($p < .0001$). Fisher's z' to r transformation indicates that the C3 correlation is significantly higher than the Rosenholtz correlation ($z = 2.3$, $p = .02$), suggesting that C3 is a stronger predictor of subjective ratings for these displays.

There was one prominent outlier in the C3-versus-ratings plot—Chart Z17 (Figure 7)—which participants rated as much less cluttered than C3 predicted (3.2 rating vs. 9.0 C3). The high C3 prediction may have been caused by repeating squiggle patterns in this chart, which generated low color density values. Participants may have interpreted the patterned areas as less cluttered than C3 could discern. Future enhancements to C3 might incorporate image segmentation or another pattern analysis technique (such as subband entropy, suggested in Rosenholtz, Li, & Nakano, 2007) to account for lower perceptions of clutter in regions with regular, repeating patterns.

Discussion

This experiment replicated and extended results from Experiment 1 and validated the C3

parameters (I and G) set in Experiment 1, which suggests that these specific parameters have a broad range of applicability. Experiment 2 demonstrated an excellent fit between the empirical subjective ratings data and both C3 and Rosenholtz clutter models.

Both the C3 and Rosenholtz models were better predictors of subjective clutter ratings for graphics in Experiment 2 than for those in Experiment 1. This may be explained partly by the fact that most of the graphics in the first experiment (unlike the second) consisted of a single, dominant feature type, the number of which appears to have influenced subjective clutter ratings significantly, as discussed earlier.

In contrast, the aeronautical charts in the second experiment comprised a much greater diversity of cartographic feature types, which were more difficult to quantify into a single feature set size. We suggest that in the case of these more complex images, in which features are not easily "countable" and set size is not obvious, clutter models such as C3 and Rosenholtz's model may be better predictors of subjective clutter (Beck et al., 2008).

CONCLUSIONS

The C3 model estimates the amount of perceivable clutter in complex geospatial displays as a function of color density and saliency. In general, low color density plus high saliency results in high clutter, whereas high color density plus low saliency results in low clutter. The model represents visually discernible features as clusters of adjacent, similarly colored pixels. Color density can be thought of as color homogeneity within features (clusters), whereas saliency is an average color difference among features.

For two different sets of geospatial displays, correlations between subjective clutter ratings and the C3 model were higher than correlations between ratings and another recently published clutter model (Rosenholtz et al., 2005). This difference in correlations was significant for the second set of displays, which consisted of aeronautical charts. Analysis of one outlier chart (Z17, in Figure 7) suggests that improvements to C3 might be gained by incorporating an image segmentation component to the

model, such that regularly occurring patterns are interpreted as less cluttered than jarring, irregular effects.

The first data set consisted of images with easily observable feature set sizes (countable numbers of similar features), which accounted for 33% of the variance in subjective clutter ratings. The second data set consisted of images with greater cartographic variability and no obvious feature set size. In both cases, C3 reliably predicted subjective clutter perception. The correlation between C3 and ratings was higher in the second data set, in which feature set size was irrelevant. We suggest that for more complex displays, in which set size is difficult to determine, C3 may be a better predictor of subjective perceived clutter (Beck et al., 2008).

The ability to predict subjective impressions of display clutter is a good first step in determining whether clutter detracts from a display's usefulness, but the ultimate test is comparing the clutter model with target search performance. Another set of experiments completed by the authors demonstrated that C3 is also a good predictor of human response time when locating targets in displays of varying clutter (Beck et al., 2008).

ACKNOWLEDGMENTS

This work was sponsored by the Naval Research Laboratory 6.2 Base Program (P.E.602435N) awarded to the first author. We thank Geary Layne and Stephanie Myrick for their important contributions. We thank the Psychology Departments at George Mason University (GMU) and Louisiana State University (LSU) for providing resources and students to participate in experiments, and Mirka Vomela (GMU) and Maghan Ferguson (LSU) for their assistance. The views and conclusions expressed in this document do not represent official policies of the U.S. Navy.

REFERENCES

- Beck, M. R., Lohrenz, M. C., Trafton, J. G., & Gendron, M. L. (2008). The role of local and global clutter in visual search [Abstract]. *Journal of Vision*, 8(6), 1171.
- Beck, M. R., & Trafton, J. G. (2007). Local spatial layout consistency affects strategies but not memory during visual search [Abstract]. *Journal of Vision*, 7(9), 1060.
- Bravo, M. J., & Farid, H. (2004a). Recognizing and segmenting objects in clutter. *Vision Research*, 44, 385-396.

- Bravo, M. J., & Farid, H. (2004b). Search for a category target in clutter. *Perception*, 33(6), 643–652.
- Commission Internationale de l'Éclairage. (2001). *Improvement to industrial colour-difference evaluation* (CIE Publication No. 142-2001). Vienna, Austria: Central Bureau of the Commission Internationale de l'Éclairage.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433–458.
- Foley, J. D., & Van Dam, A. (1984). *Fundamentals of interactive computer graphics*. Reading, MA: Addison-Wesley.
- Gendron, M. L., Layne, G. J., & Lohrenz, M. C. (2005). A system, method and apparatus for clustering features (Navy Patent Case 84,921). Stennis Space Center, MS: Naval Research Laboratory.
- Gendron, M. L., & Lohrenz, M. C. (in press). A 2-D clustering algorithm to aid mine-warfare analysts in determining navigable lanes over cluttered seafloors. *Journal of Management and Engineering Integration*.
- Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality. *Philosophical Transactions of the Royal Society*, 115, 513–583.
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259.
- Layne, G. J., Gendron, M. L., & Lohrenz, M. C. (2008). *System and method for smoothing and compression of polyline data* [U.S. Patent 7,382,370].
- Lohrenz, M. C., & Gendron, M. L. (in press). A 3D clustering algorithm to model clutter in electronic geospatial displays. *Journal of Management and Engineering Integration*.
- Lohrenz, M. C., Myrick, S. A., Trenchard, M. E., Ruffner, J. W., & Cohan, T. (2000). Pilot preferences on vector moving-map displays. *Journal of Navigation*, 53(1), 93–113.
- Neider, M. B., & Zelinsky, G. J. (2006). Searching for camouflaged targets: Effects of target-background similarity on visual search. *Vision Research*, 46(14), 2217–2235.
- Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion: A measure of display clutter. In *Proceedings of the 2005 Conference for the Association for Computing Machinery Special Interest Group on Computer-Human Interaction* (pp. 761–770). Portland, OR: Association for Computing Machinery.
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, 7(2), 1–22.
- Schons, V., & Wickens, C. D. (1993). Visual separation and information access in aircraft display layout (No. ARL-93-7/NASA-A31-93-1). Savoy: University of Illinois Aviation Research Laboratory.
- Sharma, G. (2003). *Digital color imaging handbook*. Boca Raton, FL: CRC.
- Tullis, T. S. (1981). An evaluation of alphanumeric, graphic, and color information displays. *Human Factors*, 23, 541–550.
- Tullis, T. S. (1983). The formatting of alphanumeric displays: a review and analysis. *Human Factors*, 25, 657–682.
- Waliszewski, P., & Konarski, J. (2005). A mystery of the Gompertz function. In *Fractals in biology and medicine* (Vol. 4, pp. 277–286). Basel, Switzerland: Birkhauser Verlag.
- Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, 37, 473–494.
- Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, 9(1), 33–39.
- Zuschlag, M. (2004, March). *Quantification of visual clutter using a computational model of human perception: An application for head-up displays*. Paper presented at the Human Performance, Situation Awareness and Automation II Conference, Daytona Beach, FL.
- Maura Lohrenz is a physical scientist and team leader with the Naval Research Laboratory (NRL) at Stennis Space Center (SSC), Mississippi. She received her MS in aeronautics and astronautics from the Massachusetts Institute of Technology in Cambridge, Massachusetts, in 2003.
- J. Gregory Trafton is a computer scientist, engineering psychologist, and head of the Intelligent Systems Section at NRL in Washington, D.C. He received his PhD in psychology from Princeton University in Princeton, New Jersey, in 1994.
- Melissa Beck is an assistant professor of cognitive and developmental psychology in the Louisiana State University Psychology Department in Baton Rouge, Louisiana. She received her PhD in cognitive experimental psychology from Kent State University in Kent, Ohio, in 2003.
- Marlin Gendron is a computer scientist at NRL-SSC, Mississippi. He received his PhD in applied science and engineering from the University of New Orleans, Louisiana, in 2004.

Date received: June 25, 2008

Date accepted: January 31, 2009