The Questionable Effect of Retinal Variables on Information Displays: Implications for Problem Solving and Learning

By: William A. Kealy and Chitra Subramaniam

Kealy, W. A., & Subramaniam, C. (2005). The questionable effect of retinal variables on information displays: Implications for problem solving and learning. *Journal of Visual Literacy*, 25(1), 1-20.

Abstract:
This study challenged the long-held assumption that the "retinal variable," such as color or shape, used to represent data on an information display influences how well it is interpreted and understood. Forty-eight undergraduates solved problem scenarios by consulting a multivariate display using either color, value, or shape to represent age groups of Indian tourists. The type of retinal variables employed affected neither the accuracy nor latency of question responses, regardless of whether or not the display was in view during testing. However, problem-solving tasks were rated significantly easier to perform when display information was depicted through variation in color versus value.

Article:
Designers of information displays commonly use the horizontal and vertical dimensions of a page to represent two types of variability in a criterion measure. The vertical position of a symbol or mark on the page corresponds to different levels of one factor while its location on the horizontal dimension stands for the other factor. Once a data set exceeds two factors, however, it is no longer possible to depict the information using solely the position of data points on a two-dimensional plane. In such instances, designers must resort to using what Jacques Bertin (1983) called "retinal variables," such as size, brightness (herein referred to as value), texture, color, orientation, and shape to depict variability in the additional factor.

Bertin claimed that not all retinal variables are equal in their ability to represent information—some, he contended, are more effective than others based on their capacity for portraying certain classes of data: associative, selective, ordered, and quantitative (see Figure 1). The retinal variable size, for example, is suitable for depicting ratio-level (i.e., quantitative) data such as differences in temperature because doubling a symbol's area logically signals a twofold increase in heat. Icon shape, on the other hand, is not an effective retinal variable for this type of data since there is nothing inherent about shape that graphically implies quantitative, or even ordinal, relationships. By contrast, it is fitting to represent the attributes of constructs such as gender that reflect a nominal scale of measurement (which may be further broken into Bertin's "associative" and "selective" levels) through variations in either the size or shape of symbols on a data display. Hence, one could logically argue that the effectiveness of a data display is determined, in part, by the representational capacity of the retinal variable(s) employed for portraying a particular level of data.

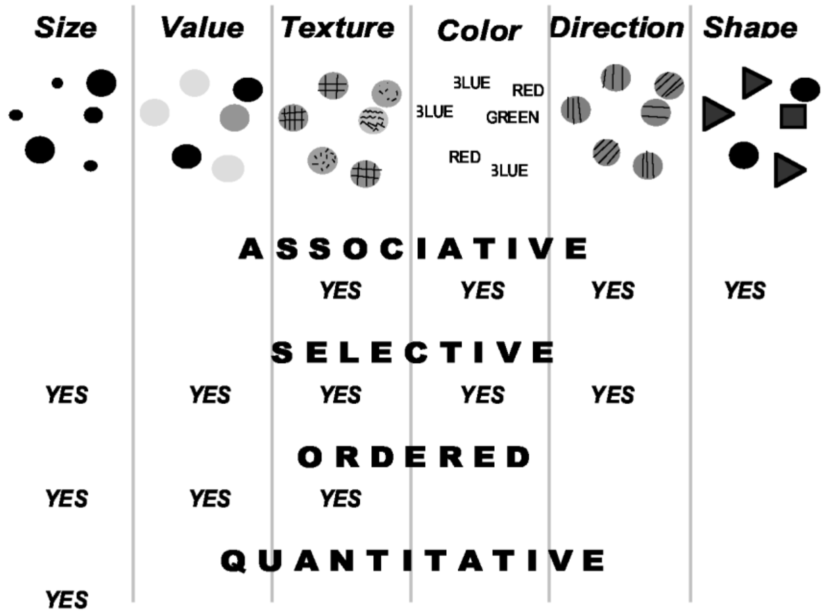

Figure 1. **Bertin's taxonomy on the relative effectiveness of various display characteristics for representing different types of display data.**

In recent years, Bertin's concept of retinal variables has received increased attention by researchers studying instructional displays like general reference maps (Griffin & Robinson, 2000; Zambo, Vogel, Nunez, & Bartshe, 2002; Verdi & Kulhavy, 2002), thematic maps (Rittschof, Griffin, & Custer, 1998), and information graphs (Shah & Hoeffner, 2002). This article, however, raises several concerns about the legitimacy of such references to Bertin's work.

First, there is a widespread misunderstanding among researchers about the notion of retinal variables. Some studies (Griffin & Robinson, 2000; Verdi & Kulhavy, 2002), for example, have misconstrued retinal variables as a type of taxonomy or checklist for map feature attributes. On the contrary, the concept of retinal variables is rooted in the study of data representation and visualization with the goal of enhancing the kinds of decisions one could make from a given data set. It is a theory that arose from an interest in improving the problem solving potential of information displays by expanding the available number of strategies for systematically portraying data.

A second criticism regarding the notion of retinal variables is its lack of a conceptual link to existing theories of cognitive information processing. Bertin neither claims that some retinal variables support better mental processing than others do nor, in the case of maps, that they are traits "that map learners can use to help identify and process the features found on the map" (Verdi & Kulhavy, 2002, p 30). Therefore, from a purely psychological point of view, there is no rationale to support the idea that better text recall results when features of an accompanying map are differentiated by using a "better" retinal variable, such as shades of gray, versus a "poorer" retinal variable like color or shape. Nevertheless, in recent years retinal variables have found their place in the studies of behavioral scientists as a component of an experimental treatment or as a criterion measure in different types of cognitive tasks.

An experiment by Rittschof, Griffin, & Custer (1998) illustrates how the concept of retinal variables has been represented in research on learning from adjunct graphic displays. In the study, participants examined one of four displays depicting ordinal data:

- a chart,
- a (choropleth) map using shades of gray to portray this information,

- a map depicting the data as circles varying in size, or
- a map with geographically-positioned numerical data.

Participants studied the display for three minutes and then read a brief related text passage for four minutes. Among participants viewing the three map versions, no differences emerged in their ability to label a blank map, recall facts from the accompanying text, or to draw inferences from what they read. The fact that no differences in performance arose from the use of maps employing different retinal variables (i.e., size and value) is not surprising since both are suitable for portraying ordinal data. Moreover, the researchers made no prediction regarding performance differences due to the use of different retinal variables. This exemplifies how, on the one hand, the concept of retinal variables has become an important consideration in research while, on the other hand, remaining unconnected to a theoretical foundation that can direct its use in experimental studies.

A third, and perhaps most critical, reason for questioning the impact of retinal variables on learning from adjunct displays is the lack of empirical evidence to support this concept. Instead, the notion of retinal variables is based solely on Bertin's introspection (Green, 1998; Nelson, 2000). Indeed, Rittschof et al (1998) point out that when cartographers and other graphic experts prescribe symbolic representation schemes to improve the clarity, comprehension, and usefulness of displays, they often do so based on their intuition. Hence, it may be fitting to question the validity of some longstanding design principles such as assumptions about the effect and use of color in instructional media (Misanchuk & Schwier, 1995).

Accordingly, the study reported herein examined the impact of retinal variables on the effectiveness of an information display used for problem solving and learning. The insights gleaned from his inquiry should be especially important for instructional designers, media specialists, and teachers who rely on such basic research when exploring ways to improve instructional graphic displays.

Theoretical Framework

According to Bertin, in some cases, there are several, not just one, retinal variables that are suitable for depicting a particular level of data. As Figure 1 illustrates, it is fitting to represent ordinal data using the retinal variables of size, value, or texture. Even so, some retinal variables appear to have a better capacity than others do for representing certain kinds of data. Logic suggests, for instance, that a representational scheme, such as one using variations in size or value, that is capable of showing an infinite number of gradations is preferable when depicting continuous a variable like temperature. By contrast, it makes less sense to use, for example, changes in shape (e.g., a circle, square, and triangle) to represent data that are continuous in nature. This, however, raises the question that is the focus of the current study: "what are the consequences for information processing and learning when a representational scheme is incongruent with the underlying characteristics of what is being represented?"

One approach to understanding the psychological role retinal variables play in the processing of information displays is to consider them as components of a symbol system (Salomon, 1994). In visually based symbol systems, the visible marks are symbolic elements that correlate with quantities or qualities existing in some field of reference. The annotations on a sheet of music, for example, are symbolic elements that correspond to specific sounds. For this correspondence to take place, the symbol system must have a scheme for doing so that incorporates a systematic method of representing elements as well as the syntactic rules for their arrangement. Returning to the example of the symbol system for music notation, the retinal variable of shape is used in the symbol scheme to depict sounds of varying durations. Consequently, retinal variables constitute an important part of the scheme for a symbol system.

Interpretation of a display, whether used alone or in concert with an accompanying text, involve both "bottom-up" and "top-down" processes, the latter driven primarily by the expected task and the type of information one anticipates having to extract from the coded message (Salomon, 1994; Schnotz, 2002). The bottom-up processes, by contrast, are mental operations evoked by the emergent characteristics of the symbol system itself including, but not limited to, its capacity for making discriminations, comparisons, and inferences. In Figure 2, for instance, it is much easier to discriminate between age categories than it is to visually isolate just the symbols representing high or low income. In a pure sense, bottom-up processes are preattentive in nature, involving the immediate organization of a visual field that precedes a consciously directed search of the information presented. Though these processes are beyond conscious control, it is nevertheless possible to consider them in the design of information displays. Healey, Booth, and Enns (1996), for example, demonstrated how a data array using features that "pop-out" during visual search (i.e., the color and direction of data points) facilitated a high-speed visual estimation of salmon migration patterns.

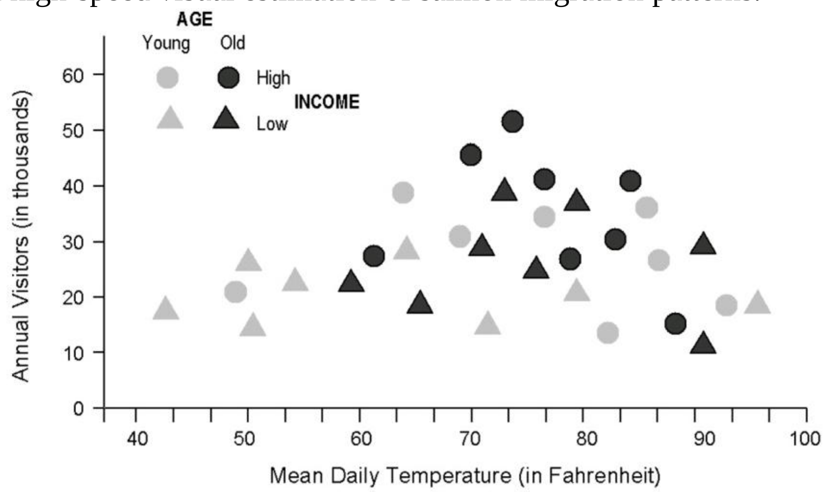

Figure 2. **Relative effectiveness of value over shape as a representational scheme for discriminating between two collocated categories of data.**

Initially, while one is learning how to interpret and "read" the scheme of an information display, these bottom-up processes predominate. However, as these decoding skills grow increasingly automatic over time, the mental activities involved in using a display becomes correspondingly top-down in nature, with task demands guiding attention allocation (including some bottom-up processes), knowledge building, and higher-order thinking such as hypothesis testing. Salomon (1994) proposes that an improvement in both the usefulness of a multivariate display and the learning that results from such use occurs "when the 'top-down' processes (determined by the task) are compatible with the 'bottom-up' processes (determined by the symbol system)" (p. 226).

Bertin (1983) provided what is essentially a bottom-up explanation of display processing that involves identifying a) the variables represented, b) the method used for representing the variables, and c) how the interaction of variables is shown on the display. Theoretically, use of the inappropriate retinal variable in a display may inhibit its processing, as Figure 2 illustrates in its use of shape—a retinal variable that Bertin asserts does not have the capacity for evoking selectivity when used to depict data (see Figure 1).

By contrast, Wainer (1992) proposed that the reading of information displays such as graphs and tables is a task-driven (i.e., top-down) process involving three levels of mental processing: a) retrieval of isolated facts, b) identification of local relationships, and c) recognition of global trends suggested by the data. While not intended to correspond to either ordered stages in the reading of a graph or increasing degrees of mental difficulty, these levels nevertheless suggest a hierarchy in which data portrayal at the elementary level is a prerequisite for more sophisticated data representation. Two or more data points on a graph (level 1), for instance, enable percipients to make comparative judgments (level 2) while sets of related comparisons yield the capacity for making inferences about the data (level 3).

Whether one views graphic displays in terms of their processing stages (Bertin) or their processing capacities (Wainer), the distinction naturally raises the question of whether the influence of retinal variables is the same regardless of the level at which a display is read. When learners make an effort to understand (or deeply process) a text, they see beyond its surface features, attending instead to the semantic nature of what they are reading (Craik & Lockhart, 1972). Similarly, when learners exert mental effort in using a data display to solve a problem, the specific retinal variable employed may be less critical than Bertin's model would suggest.

Alternatively, a "retinal variable effect" may only be evident in circumstances demanding a more complex reading of a display—situations where percipients must retain several display variables concurrently in working memory. Further, it is conceivable that, while not affecting cognitive processes, choice of the particular retinal variable(s) used in a chart, diagram, or map could influence processing motivation through the "perceived mental effort" (Salomon, 1984) that viewers imagine is required to derive meaning from the display.

To address this possibility, we constructed data retrieval tasks requiring increasingly complex visual scrutiny of the data display. We reasoned that violating Bertin's prescription for representing an ordinal variable on a display (e.g., use of shape to depict age categories) might have a detrimental effect only when a paucity of search criteria demanded a more extensive scan of the display.

Method

Design and Participants

Forty-eight undergraduate education majors took part in the study, receiving credit on their course grade for participating. Participants viewed displays using either shape, value, or color to depict information. Displays used to solve three types of scenario problems varied in the complexity of display scanning required to find a solution. Hence, the design for the study was a 3 Display (Shape vs. Value vs. Color) x 3 Scan (Low vs. Medium vs. High) factorial with Display varied between participants and Scan serving as a repeated measure. Dependent variables included accuracy and latency measures on both a data search task (with display present) and a comparative judgment task (with display absent).

Materials

Displays

For the experiment, we created a data graphic, displayed on a standard 15-inch (38.1 cm) monitor (set to 800 x 600 pixel SVGA mode), on the rates of tourism to five Indian cities—a topic specifically selected for its likely unfamiliarity among study participants. All displays depicted four variables: City (Simla, Panaji, Delhi, Bombay, and Chennai), Season (spring, summer, fall, winter), Income (above or below 50,000 rupees per year), and Age Group (20-35, 35-50, 50-65). City labels, in dark gray 10-point Arial Black lettering, were vertically stacked (in the same order, top to bottom, as previously stated) in the center of a white computer screen with season labels printed in black 9-point Arial letters directly below each city (see Figure 3). Light gray (RGB value = 206, 207, 206) horizontal guidelines extended from the left and right of each season label to about 60 pixels from the respective edges of the screen. Three icons distributed along these lines to the left of every season label indicated tourism rates by travelers in different age groups earning less than 50,000 rupees annually. A similar arrangement of icons to the right of seasonal labels showed tourism rates among those with incomes over 50,000 rupees. This formed two back-to-back "stem-and-leaf" plots (Tufte, 1983) in which the number of tourists per week, indicated by an icon's horizontal distance from the center, ranged from zero to 6,000 (at the left and right edges). To aid readability, light gray vertical lines marked off increments of 1000 tourists.

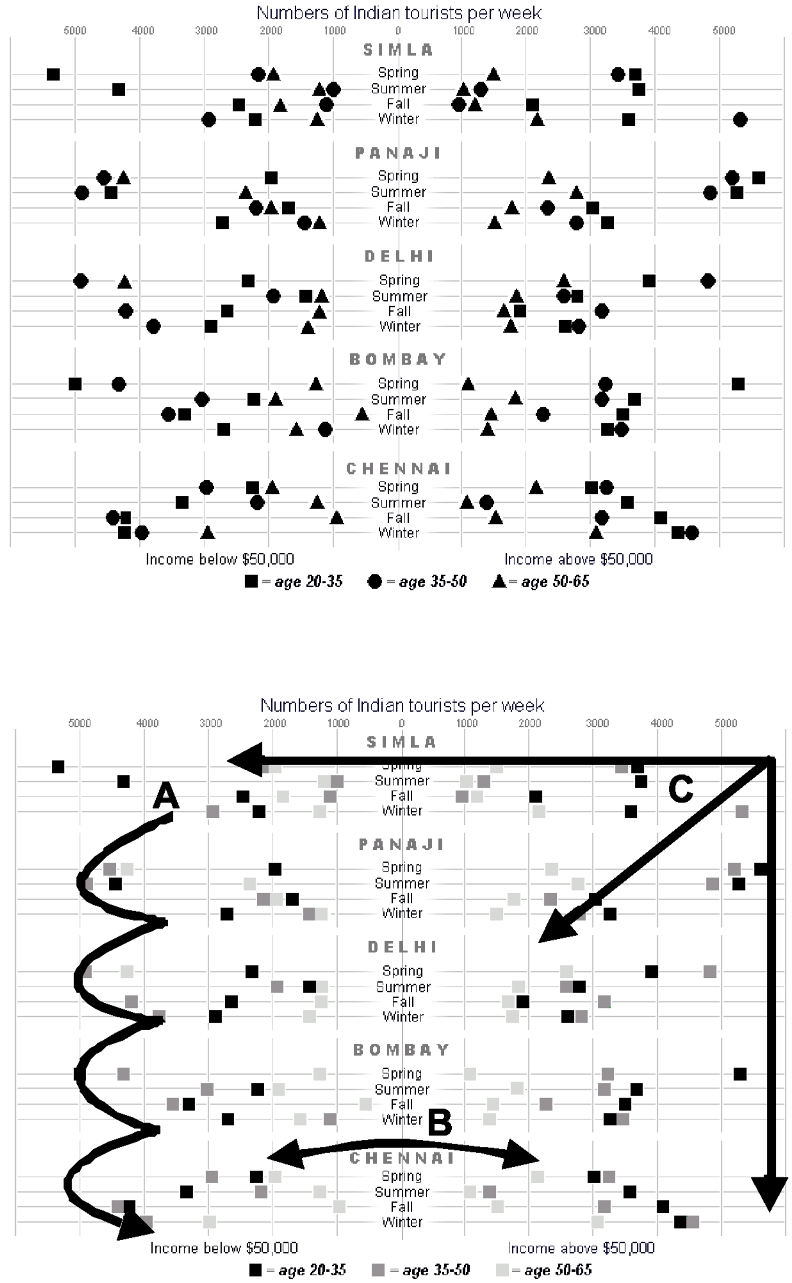

Figure 3. The Shape and Value experimental displays (shown top and bottom, respectively) used in the. Annotations on the Value display indicate the patterns of data search necessary for answering A) Low Scan questions, B) Medium Scan questions, and C) High Scan questions.

For the Shape experimental condition, displays depicted the three age groups of tourists (20-35, 35-50, and 50-65 years old) using squares, circles, and equilateral triangles, respectively (see Figure 3). In the Value condition, the same age groups were respectively portrayed using black, dark gray, and light gray squares (RGB components of 0–0–0, 160–160–160, and 220–220– 220, respectively). Finally, in the Color experimental condition the three age groups were represented using blue (ages 20-35), green (ages 35-50), and red (ages 50-65) squares. To guide our color selection, we chose ones with hues that had the greatest perceptual discriminability from one another and, at the same time, virtually identical in brightness (Healey, Booth, & Enns, 1996). The RGB components of the blue, green, and red colors were, respectively, 76–76–255, 0–166–0, and 217–0–0. In order to exercise some experimental control for the non-critical aspects of display design, we constructed alternate versions of the three displays by changing both the symbol used to depict each age group and the order for listing the cities; this yielded the six types of displays using in the study. For example, in the alternate version of the Color display, red, green, and blue respectively represented the 20-35, 35-50, and 50-65 age groups while the cities listed were, from top to bottom, Bombay, Delhi, Simla, Chennai, and Panaji.

Scenarios

The study employed three varieties of scenario-like questions, six in each category, (18 questions total) that varied in the complexity of visual scan required to find a solution using the display. Specifically, we defined scan complexity in terms of the number of data points that a question required participants to consult on the display. In this manner, we created three categories of questions to represent three levels of scan complexity:

1. Low Scan: Where do most senior citizens choose to take a low- budget vacation during the winter months?
2. Medium Scan: When is the least popular time for vacationing to Delhi among couples in their 40s with children?
3. High Scan: What time and location is most popular for taking a vacation among tourists in the 35-50 age bracket?

The display portrayed at the bottom of Figure 3 illustrates the visual search required for the three types of questions. In the case of Low Scan ("where") questions, participants needed to identify an ideal destination when provided three search criteria: age, income, and season of travel. To solve this scenario question, participants needed to consult and compare just five data points (shown by the arrow marked "A" in Figure 3). To answer Medium Scan, or "when" questions, participants determined the best season to travel given only location and age of travelers as the search criteria. Consequently, participants needed to examine eight data points on the display (see arrow B). By contrast, High Scan (i.e., "what") questions provided participants with only one criterion, either season or age, demanding consideration of 30 or 40 data points, respectively (arrow C). In reality, however, the number of points to consult would be much smaller since only the extreme values on the perimeter of the display are of interest.

Within each level of scan complexity, we attempted to represent the different variables of the display (age, city, income, season) in roughly equal numbers. For instance, Medium Scan questions represented the two income levels of travelers in equal numbers and mentioned each season at least once.

Participants viewed the scenario questions, data displays, and instructions through a computer program we designed and developed using the Authorware (1997) programming software.1 The program also presented a detailed 14- screen tutorial on how to read a multivariate display using a simple example on beverage consumption by men and women during lunch and dinner. The tutorial briefed participants on the three types of scenario questions they would be answering and provided practice and guided feedback on the responses made. We installed six different programs, representing two versions of the three experimental conditions, in roughly equal numbers on workstations of a large computer lab. An inactive workstation was positioned to the left and right of each programmed workstation, with every other row of computers disabled, to prevent participants from viewing treatments to which they were not assigned. For all computer workstations used in the study, we inserted a standard 3.5-inch (8.9 cm) floppy disk before each session to record response data from each participant.

Procedures

Upon arrival at an experimental session, the researcher randomly assigned each participant to one of the programmed computers, which, at the time, showed blank gray-colored screens. This random assignment resulted in the following distribution of participants to the three treatment groups: Color- 15, Value-17, Shape-16. Reading from a prepared script, one of the experimenters welcomed participants, informing them that for the next 30- 40 minutes they would be using a data display to answer scenario type questions. Additionally, they would also be answering two-choice scenario questions using whatever data they could recall from the display. The experimenter then reminded participants that they could quit the study at any time without penalty and that if a question or computer malfunction arose during the session, they should raise their hand for assistance.

Once all procedural questions were answered, participants pressed the TAB key and followed the on-screen instructions printed in dark gray 18- point Arial typeface on a light gray background. The first screen gave an overview of the task participants would be performing. By clicking a button marked "click to continue" at the bottom of the screen, participants advanced to the second screen that stated they had the right to quit the study at any time without penalty and were to raise their hand for assistance in the event of a computer malfunction or if a procedural question arose. The last sentence of the screen asked participants to raise a hand if any questions remained. Once it was clear to the experimenter that there were none, participants clicked on the word "away" in the text, branching them to the next screen.

Participants next saw a 14-screen tutorial that explained how to read and interpret the data display that they would be using. The tutorial also provided them with practice questions that were similar in style to the ones answered later in the session. The first screen showed a data display on the preference for four different types of beverage during lunch and dinner among men and women in three age groups. Though simpler than the display on tourism viewed later, the sample was identical to the experimental display in graphic layout, number of variables depicted, and retinal variables used. Forward and back arrows, respectively located in the lower right and lower left corners of the screen permitted participants to proceed through the tutorial at their own pace.

The second screen of the tutorial introduced the concept of "reading" a data display while the third screen presented a one type of question in which three variable were provided (i.e., a Low Scan question). Orange lines that annotated the display on this screen revealed how to read it to determine the correct answer. The fourth screen provided the answer and an example of a question with two variables provided (i.e., a Medium Scan question). The screen that followed illustrated how to read the data display for this type of question and the sixth screen gave the answer and a detailed explanation. On the seventh screen of the tutorial, participants viewed the last type of question consisting of just one variable to direct the search (i.e., a High Scan question) and orange circles identifying the many data points that one needed to scan to find the answer. Participants saw, on the next screen, the answer to this question and why it was the correct choice, the ninth screen presented them with an alternative version for this type of question, and the tenth screen gave the solution. The tutorial's eleventh screen instructed participants to answer each scenario question by clicking the mouse on the display symbol providing the data for the problem solution. The next tutorial screen stated that, for correct responses, the word "correct" would appear in the upper right corner. Screen thirteen of the tutorial illustrated that an incorrect response would yield both the printing of the word "incorrect" in the upper right corner and the highlighting, in orange, of the symbol that represented the correct data point.

The last screen of the tutorial told participants they would have an opportunity to practice sample questions using the data display on beverage consumption. Participants began the practice by clicking the TAB key and responded to four questions reflecting each type of question—Low, Medium, and High Scan (both versions of a High Scan question)—presented in random order. Following both correct and incorrect responses to each question, detailed feedback explained how the display provided the solution to the problem.

After completing the last practice problem, participants viewed a screen informing them that they would now view a display on rates of tourism to five Indian cities by travelers in three age groups. After a two-minute study period, timed by a clock-shaped icon on the screen, they would use the display to complete 18 scenario questions similar in form to the questions previously encountered. The participants then looked up at the experimenter to signal that they had finished reading the instruction, raising their hand if there was a question about the procedure described. Participants were then told to click on the number "50,000" in the first paragraph of the onscreen text to reveal the display on tourism. At the end of the two-minute period, participants saw a screen of instructions telling them they would now answer the 18 scenario questions by clicking the single data icon on the display that best answered the problem presented. The screen also told them that, following each response, the computer would provide feedback on its correctness by printing the word "correct" or "incorrect" at the top of the screen and, if incorrect, highlighting the correct answer in orange. To then view the next question, participants were to press the TAB key. Finally, the directions told participants to do the best they could in answering each question while working as quickly as possible. Participants then pressed the TAB key to view the first of 18 questions presented in random order. For each question, the computer program recorded both the accuracy and latency of response in seconds (to the third decimal place).

Immediately after the last question was finished, a screen appeared telling participants that in a moment they would answer the same questions about travel in India completed earlier, but without the aid of the display. In this case, however, they would respond by clicking on one of two rectangles located at the left or right half of the screen near the bottom. The screen presented participants with a sample question using data from the display on beverage consumption and two choices shown as darker gray rectangles that occupied the left and right halves of the bottom one-fifth of the screen. Centered in one of the rectangles was the correct answer while the other contained an incorrect answer, both printed in black 18-point Arial Bold typeface. When clicked, the rectangle chosen flashed in inverse and another sample question appeared followed by a set of instructions informing participants that a blank screen with an asterisk in its center would precede each question for a period of two seconds. Participants next clicked the TAB key, completed two practice questions dealing with beverage consumption, and read a screen of instructions stating they would now answer 18 questions similar to the ones just completed, but dealing with the tourism data studied earlier. The onscreen text urged participants to do the best they could at answering each question while responding as quickly as possible. They then pressed the TAB key and answered the 18 two-choice questions, presented one-by-one in random order.

As soon as participants answered the last question, a new screen asked them to rate how easy or hard it was to accomplish the tasks in the session by clicking one of the buttons located at the bottom of the screen. Between the text and the buttons appeared the question, “How easy or hard was it to answer the questions using the data display provided?” printed in a blue 18-point Arial Italic typeface. Below this were 10 square buttons, numbered 1 to 10, arranged in a straight row across the screen. Immediately to the left of the first button and to the right of the last button appeared, respectively, the words “easy” and “hard” printed in blue 18-point Arial Italic capital letters. Once participants clicked on a button to make their rating, the question was replaced by one asking, “How easy or hard was it to answer the questions from memory?” Participants then completed the second rating, read two screens briefing them on the nature of the study and thanking them for their participation, clicked a button to log off the computer, and were dismissed from the room. An experimental session typically lasted 30-40 minutes.

Results

All tests of significance used .05 as an alpha level. Initially, we conducted a boxplot analysis of performance scores and latencies for data finding and data recall tasks. This revealed three participants, one in each treatment group, exhibiting three or more measures characterized as either an outlier or an extreme scores: values, respectively, either between 1.5 and 3 times the interquartile range (i.e., the distribution of scores between the 75th and 25th percentile) or more than 3 times this range, from either end of the range. One participant in the Shape treatment group, for example, produced outlying scores for the Medium and Low Scan questions dealing with data finding and outlying latencies for Low and High recall questions. The removal of these three participants from the subsequent analysis resulted in sample sizes of 15, 16, and 14 for the Shape, Value, and Color conditions, respectively.

Differences due to the two versions of program used for each of the three treatments were examined with a 3 Display x 2 Version ANOVA. The analysis did not indicate significant main effects or interactions for any of the four measures of cognitive performance. Consequently, we pooled these data, eliminating Version as a factor for the subsequent analysis. The mean percentage of correct responses and mean response latencies for scenario questions answered with and without the data display in view by participants in the three treatment groups are shown in Table 1.

Table 1

Accuracy and Latency of Response for Finding and Recalling Data on Multivariate Displays Using Shape, Value, or Color to Depict Information.

Display Condition		Type of Scan Imposed by Questions Low	Medium	High
Shape	Data search (%)	.70 (.24) [a]	.83 (.15)	.86 (.17)
	Latency (sec)	27.39 (7.01)	27.21(5.56)	32.94 (9.90)
	Data Recall (%)	.49 (.28)	.63 (.22)	.42 (.23)
	Latency (sec)	6.67 (1.78)	6.62 (1.64)	*8.67 (2.11) [b]
Value	Data search (%)	.70 (.25)	.73 (.28)	.79 (.25)
	Latency (sec)	28.49 (7.42)	28.53 (6.13)	32.53 (9.04)
	Data Recall (%)	.38 (.20)	*.55 (.15)	.44 (.19)
	Latency (sec)	6.39 (1.84)	6.67 (1.93)	*9.38 (2.67)
Color	Data search (%)	.69 (.23)	.81 (.18)	*.89 (.11)
	Latency (sec)	27.48 (8.51)	27.65 (7.98)	31.83 (10.13)
	Data Recall (%)	.36 (.24)	.51 (.24)	.55 (.28)
	Latency (sec)	7.14 (2.77)	6.28 (3.31)	8.98 (3.00)

Note: n = 15, 16, and 14 for Shape, Value, and Color treatment groups, respectively.

[a] Standard deviations for all mean percentages are shown in parentheses.

[b] Asterisks indicate a significantly different value on univariate tests for scan type.

Use of Data Displays

Mean percentage correct scores on scenario questions answered using the data displays were 80, 74, and 80, respectively, for participants using Shape, Value, and Color displays. Results of a 3 Display x 3 Scan repeated measures ANOVA indicated none of these differences were significant. However, the main effect for Scan, $F(2, 84) = 7.51$, $p = .001$, $d = .94$, revealed participants differed significantly in the accuracy of their responses due to the type of question presented, which accounted for about 15% of the variability in scores. A Bonferroni post hoc analysis revealed participants performed significantly better on High Scan questions $F(2, 132) = 5.54$, $p = .005$, where one search criterion was provided, compared to Low Scan questions in which three criteria were given.

Mean latency of response by participants in the Shape, Value, and Color experimental conditions was virtually identical (29, 30, and 29s, respectively). As expected, a 3 Display x 3 Scan repeated measures ANOVA showed no differences in the time participants took to respond to scenario questions attributable to differences in the displays they used. By contrast, a significant main effect for Scan, $F(2, 84) = 12.13$, $p < .001$, $d = .99$, indicated participants varied greatly in their latency of responses as a result of the type of question they answered. On the average, High Scan questions took over four seconds longer to answer than either Medium or Low Scan items.

Recall from Data Displays

Without the display in view, the proportion of correct responses fell to .51, .46, and .47, respectively, among participants in the Shape, Value, and Color group. A 3 Display x 3 Scan repeated measures ANOVA revealed that neither the main effect for Display nor the Display x Scan interaction to be significant. On the contrary, a significant main effect was observed for Scan for both response accuracy, $F(2, 132) = 5.56$, $p = .005$, and latency, $F(2, 84) = 48.33$, $p < .001$, $d = 1.0$. Univariate analyses and follow-up Bonferroni test showed revealed significantly higher recall performance on Medium questions relative to Low questions, $F(2, 141) = 4.97$, $p = .008$, $d = 1.0$, while High Scan questions were answered significantly slower—by over two seconds— than either Low or Medium Scan questions, $F(2, 132) = 15.65$, $p <.001$.

Perceived Task Difficulty

To analyze the perceived difficulty among the three groups of participants for answering scenario questions with the display either present or absent, we entered the rating data in a 3 Display x 2 Task (display present or absent while responding to a scenario question) repeated measures ANOVA. The analysis indicated, as expected, significantly higher perceived difficulty in answering questions without the aid of the display (i.e., solely from recall of the data studied) compared to when the display was in full view, $F(1, 39) = 185.65$, $p < .001$, $d = 1.0$. Although, the main effect for Display showed no significant global differences in difficulty ratings based on the type of display used, the analysis reported a significant, $F(2, 39) = 4.83$, $p = .013$, $d = .77$, Display x Task interaction. Figure 4 graphically depicts this, illustrating that between-subjects differences in ratings emerged only for the task in which the data display was present. Follow up univariate tests for the simple effects of Display at each level of Task indicated significant, $F(2, 40) = 3.13$, $p = .05$, differences between the treatment groups regarding the perceived difficulty of answering questions when the display was present. A Duncan post hoc test showed the Color group (M=2.36) rated the task significantly easier than did the Value (M=4.43) group. Ratings by the Shape group (M=4.13), however, did not differ significantly compared to either of the other groups. None of the groups differed from one another in their ratings on the difficulty of completing scenario questions when the data display was not available.

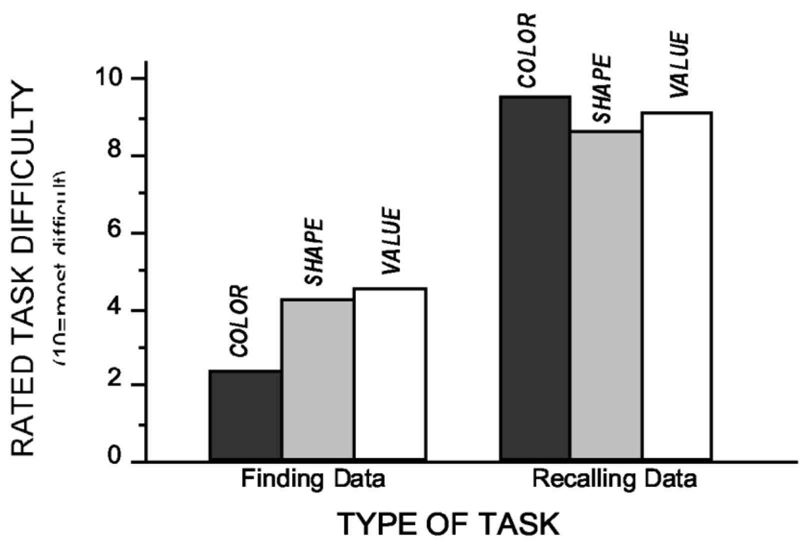

Figure 4. **Participant ratings on the perceived difficulty in answering scenario questions with data displays, rendered using different retinal variables, that were either present or absent at the time of testing.**

Discussion

A primary purpose of this study was to examine the validity of Bertin's retinal variables as a factor in the accessibility of information from data displays. In general, the study found little support for the notion that the retinal variable used in a data display influences how quickly or accurately people use it to solve a problem. In terms of bottom-up processing, we expected a "retinal variable effect" to be reflected in the speed and accuracy of a data search, across all levels of task difficulty, for one display compared to another. Although use of the Value display resulted in data retrieval that was both less accurate and longer in duration than the Color and Shape displays, these differences were not statistically significant.

From the perspective of task-driven or top-down processing, however, there is evidence, albeit tenuous, that the particular scheme used to represent data may exert an influence over how effectively a display is used. For all between-subjects groups, accuracy for answering scenario questions, as well as the time to respond to them, increased as the number of variables provided to aid data search decreased. Across treatment conditions, responses to High Scan questions were significantly more accurate than Low Scan questions, though at the expense of significantly greater processing time—conceivably, this is because a relatively smaller set of search criteria provokes a more focused exploration of a display. Within each treatment condition, however, these task-related differences in performance were only significant for participants with displays using the retinal variable of color. Though highly speculative, it may be possible that use of color yields better visual discrimination of data compared to value or shape, but that this advantage is only realized for tasks involving the scrutiny of a large number of data points.

A secondary aim of the experiment was to determine the implications of retinal variables for the recall of information from a display. In the current study, the particular retinal variable used to portray a display's data did not influence participant's ability to solve scenario-type problems using just what they could recall from their previous study of the display. The mean percentage of correct responses on this task for each treatment group was roughly 50%, which, given the two-choice question items used and the relatively fast responses (6-9 s.) of participants, suggests their answers were essentially guesses. This possibility plus the very high ratings on the difficulty of this task—a mean of 9.14 out of 10 for all participants—underscores the difficulty they had in answering questions solely by remembering the data viewed earlier on the display. Consequently, the low scores of participants probably masked any between-subjects variability in performance that may have existed. In all likelihood, the complexity of the tourism displays themselves contributed to the difficulty in remembering the data they presented. These displays portrayed four variables: location, season, income group, and age group. By contrast, Bertin (1983) asserts that displays should not represent more than three variables (i.e., two planar variables and one retinal variable). Hence, we recommend that future studies exploring the differential effects of retinal variables on recall limit the number of variables represented on an experimental display to three.

One conclusive finding from our study was that participants considered the task of solving scenario problems significantly easier when the adjunct display they consulted rendered data through color than when shape or value served as the retinal variable. This has important implications for the effectiveness of displays since, by lowering the perceived cognitive demand associated with a display, viewers would conceivably be willing to invest a greater amount of mental effort in processing the information presented (Salomon, 1984). The consequence of design decisions, such as choice of retinal variable, on a display's perceived ease of use as well as the processing motivation of its viewers is, we believe, a potentially rich area for future research.

This study underscores the importance of an ongoing research agenda that questions longstanding but unfounded prescriptions for instructional media. In the present study we found no evidence supporting the idea that one retinal variable is better than another for depicting a given type of information (i.e., in the present case, the relative efficacy of value over either color or shape for representing interval data). While this outcome, by itself, is insufficient grounds for rejecting Bertin's concept of retinal variables altogether, it nevertheless demonstrates the need for researchers to challenge many of the established design prescriptions, rooted in professional practice, that have never been validated through empirical study. More broadly stated, a role of instructional media research is to continually question the assumptions of what constitutes valid practice in information display design and to exemplify the kind of critical judgment that remains an important aspect of media literacy (Kealy, 2004).

Notes

A fully-functioning copy of the Authorware program used for the Color treatment group may be downloaded from: http:// www.coedu.usf.edu/kealy/retinal The extension ".exe" (minus the quote marks) should be added to the filename prior to running the program.

References

Authorware 4.0.2 [Computer software]. (1997). San Francisco, CA: Macromedia.
Bertin, J. (1983). *The semiology of graphics: Diagrams, networks, maps*. Madison, WI: University of Wisconsin Press.
Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 761-784.
Green, M, (1998). *Toward a perceptual science of multidimensional data visualization: Bertin and beyond*. Retrieved July 22, 2003 from the World Wide Web: http://www.ergogero.com/dataviz/dviz0.html
Green, M. A. (1998). *Basic color and design "should be frequently asked questions" (SBFAQ)*. Retrieved December 26, 2003 from the World Wide Web: http://www.ergogero.com/FAQ/cfaqhome.html
Griffin, M. M., & Robinson, D. H. (2000). Role of mimeticism and spatiality
in textual recall. *Contemporary Educational Psychology, 25*, 125–149.
Healey, C. G., Booth, K. S., & Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interfaces, 3*(2), 107-135.
Kealy, W. A. (2004). Media Literacy. In D. Weil & J. L. Kincheloe (Eds.), Critical thinking and learning: An encyclopedia. Greenwood Publishing Group.
Misanchuk, E. R., & Schwier, R. A. (1995). The Mythology of Colour in Multimedia Screen Design: Art, Science, and Connoisseurship. *Canadian Journal of Educational Communication, 24*(1), 3-26.
Nelson, E. S. (2000). Designing effective bivariate symbols: The influence of perceptual grouping processes. *Cartography and Geographic Information Science, 27*(4), 261-278.
Rittschof, K. A., Griffin, M. M., & Custer, W. L. (1998). Learner differences affecting schemata for thematic maps. *International Journal of Instructional Media, 25(*2), 179-194.
Salomon, G. (1994). *Interaction of media, cognition, and learning: An exploration of how symbolic forms cultivate mental skills and affect knowledge acquisition* (2nd ed.). Hillsdale, NJ: Erlbaum.
Salomon, G. (1984). Television is easy and print is tough: The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology, 76*(4), 647-658
Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Réviéw, 14*(1), 101-120.
Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Réviéw, 14*(1), 47-69.
Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
Verdi, M. P., & Kulhavy, R. W. (2002). Learning with maps and texts: An overview. *Educational Psychology Réview*, *14*(1), 27-46.
Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher, 21*(1), 14-23.
Zambo, D., Vogel, J., Nunez, L., & Bartshe, P. (2002, April). *Attributes of maps for text recall*. (Mentoring for Academic Writing Roundtable Program) conducted at the Annual Meeting of the American Educational Research Association, New Orleans, LA.