1993

EYE MOVEMENT-BASED ANALYSES OF GRAPHS AND TABLES: THE NEXT GENERATION

Gerald L. Lohse
The Wharton School of the University of Pennsylvania

EYE MOVEMENT-BASED ANALYSES OF GRAPHS AND TABLES: THE NEXT GENERATION

Gerald L. Lohse
Nelson Peltz Term Assistant Professor of Decision Sciences
The Wharton School of the University of Pennsylvania

ABSTRACT

An empirical study examined the accuracy, decision speed, and eye movements of twenty subjects using color and monochrome bar graphs, line graphs and tables for performing three information acquisition tasks: point reading, comparisons, and trends. The prevailing theories from the MIS graphics literature did not predict the results. The paper advocates that the next generation of graphics research adopt a more objective theory of graphics that provides robust quantitative predictions for evaluating the quality of competing graphic designs.

1. INTRODUCTION AND THE CONCEPTUAL RESEARCH FRAMEWORK

Designers of screen displays, graphical user interfaces, multimedia displays, and electronic image processing systems need detailed, well-proven engineering principles for making graphic design decisions. Often help is embedded in graphics presentation software. Some vendors even include phone support with expert graphic designers. "Help" can also take the form of a graphics handbook or a set of design guidelines. The most prominent books on graphics (Bertin 1967; Tufte 1983) have been based on the authors' intuitions. While these intuitions have yielded valuable insights, recent work in visual psychophysics discredits some of the sweeping generalizations suggested by some graphics design guidelines (Legge, Gu, and Luebker 1989; Spence 1990; Spence and Lewandowsky 1991).

To reduce the time spent preparing graphs, tables, and text slides as well as to increase the effectiveness of the presentation, some computer science graphics research programs have automated graphic design. The *Display Analysis Program* by Tullis (1986, 1988) is a tool for testing the effectiveness of screen designs and suggesting alternative layouts. Currently, the program is limited to alphanumeric, non-graphic displays. Mackinlay (1987) developed a compositional algebra for automatically generating a wide variety of graphics from a small set of presentation elements using a computer program called APT (*A Presentation Tool*). Designs are subject to effectiveness and efficiency criteria developed by Mackinlay. Unfortunately, the lack of a quantitative theory of graphics precluded a more complete implementation. Casner and Larkin (1989) use theories from cognitive psychology about visual information processing to redesign complex airlines reservation displays using a program called BOZ. BOZ replaces demanding logical operators with less-demanding operators that reduce visual search. However, the practical application of BOZ is limited because BOZ lacks criteria for evaluating the effectiveness of a display design.

The MIS literature contains numerous studies that evaluate the effectiveness of business graphics. Experimental factors that influence performance include characteristics of the task, graphic display design factors and expertise of the decision maker (Ives 1982; DeSanctis 1984; Benbasat, Dexter, and Todd 1986; Jarvenpaa and Dickson 1988; Jarvenpaa 1989; Kleinmuntz and Schkade 1989). The MIS experimental paradigm led to the collection of data assessing the efficacy of graphic decision aids by comparing several experimental factors at a time. Unfortunately, this experimental factors approach did not allow researchers to make convincing *a priori* predictions about the expected study results. Furthermore, even when these experiments found differences among the treatments, they failed to integrate the results across studies to explain complex interactions among the experimental factors and provide objective predictions.

Recent MIS graphics research has adopted an information-processing paradigm from the cognitive psychology and behavioral decision theory literature. Jarvenpaa (1989) used the effort/accuracy theory of contingent behavior to examine how people process information from two types of bar graphs. She found that participants adopted linear or conjunctive rules to select a restaurant site when the bar graphs quantifying restaurant features were organized by each site alternative. Participants tended to use majority of confirming dimensions or elimination-by-aspect rules when bar graphs were organized by the attributes of all restaurant sites. Jarvenpaa noted that contingent decision making is a good theoretical basis for understanding the effects of information format on information acquisition behavior.

Kleinmuntz and Schkade (1989) expanded the effort/accuracy framework by examining the cognitive implications of graphic decision aids. They argued that information format predisposes the decision maker toward using a set of information processing strategies that require minimal effort given the display. Although they make specific research propositions about the impact of displays on decision strategies, their propositions are not evaluated empirically.

Vessey (1991) describes a theory of cognitive fit between tasks and graphic decision aids. The theory of cognitive fit states that graphs emphasize spatial information that can be viewed at a glance, while tables emphasize discrete data values. She also notes that some tasks assess the data as a whole (e.g., comparison and trend questions) while other tasks emphasize discrete data values (e.g., point reading questions). The theory of cognitive fit states that using graphic decision aids that fit the task increases speed and accuracy of performance. For example, tables should be better for reading individual data values than either bar graphs or line graphs, but line graphs should be better than tables for detecting trends. Her meta-analysis of the conflicting results in graphs versus tables studies provides support for the theory. Moreover, her predictions fit the guidelines recommended by Jarvenpaa and Dickson (1988) as to when and how to use business graphics. However, the theory of cognitive fit was not tested empirically.

Tan and Benbasat proposed a decomposition taxonomy to match tasks and graphical representation (1990) and conducted three experiments to compare the effectiveness of bar, line and symbol graphs for performing information acquisition tasks (in press). They focused on interactions between graph and task and found a significant interaction for decision speed but not for decision accuracy. Unfortunately, they did not include tabular displays in their study to permit a more direct evaluation of Vessey's theory of cognitive fit.

Lohse (1991a, 1991b) formalized the human information processing mechanisms for elementary graphical tasks alluded to by Benbasat, Dexter, and Todd (1986), Jarvenpaa (1989), Vessey (1991), and Kleinmuntz and Schkade (1989) in a computer program called UCIE (Understanding Cognitive Information Engineering). Using assumptions about eye movements, STM capacity and duration limitations, and the relative level of difficulty to acquire information in each glance, UCIE assigns timing parameters to most elementary graphical perception tasks (Card, Moran, and Newell 1983; Olson and Olson 1990). By summing the time required for each component task, UCIE predicts the total reaction time to answer a question posed to a bar graph, line graph or table. An empirical study compared actual performance to UCIE predictions over a range of 576 combinations of presentation formats and question types for each of twenty-eight subjects. For conditions that predominantly involve serial processing, zero-parameter predictions from UCIE explained 60% of the variance in reaction times.

The research presented in this paper is the beginning of a larger two-year program of research investigating how people process information from graphical displays and applying computational cognitive science to model these processes. While the predictions from UCIE are not the focus of this paper, the paper motivates the computational cognitive engineering approach for evaluating the effectiveness, efficiency and quality of graphic designs. The paper suggests that the next generation of graphics research adopt a quantitative focus that not only quantifies our knowledge of graphical information processes but also aids the development of metrics for evaluating the quality of competing graphic designs.

Section 2 describes a laboratory experiment that uses eye movement processing tracing data to evaluate the effectiveness of bar graphs, line graphs and tables for performing three information acquisition tasks: point reading, comparisons, and trends. Section 3 reports statistical analysis of accuracy and decision time as well as a decompositional analysis of decision time components using eye fixation data. Section 4 discusses the findings and describes the difficulty of predicting graph effectiveness for a specific task using qualitative theories. The paper concludes by suggesting a more predictive and quantitative focus in future MIS graphics research programs.

2. METHODS

2.1 Research Hypotheses

The graphs versus tables area has been a popular topic of debate in the MIS literature (Ives 1982; DeSanctis 1984; Benbasat, Dexter and Todd 1986; Jarvenpaa and Dickson 1988; Kleinmuntz and Schkade 1989). Some studies found that graphs were more effective than tables; other studies found that tables were more effective than graphs. The vanguards of the field postulated that the effectiveness of a particular graphic format was contingent upon the task. For example, Jarvenpaa and Dickson recommend using tables for reading values from the display using bar graphs for displaying comparison information, and line graphs for displaying trend information. Also, Vessey states that using graphic decision aids that fit the task increases speed and accuracy of performance. This study examines the effectiveness of bar graphs, line graphs and tables for performing three information acquisition tasks: point reading, comparisons, and trends. If the prevailing theories are correct, tables should be faster and more accurate for reading individual data values, bar graphs should be faster and more accurate for making comparisons, and line graphs should be faster and more accurate for detecting trends.

The current study also evaluates the effectiveness of monochrome and color displays; however, it is not clear how contingency theory or the theory of cognitive fit would predict the effects of color on graphic decision aids. Benbasat, Dexter, and Todd found that color displays enhanced discrimination, especially for subjects under time pressure. Color also enhances perceptual grouping (Kinney and Huey 1990). Background shading, underlining, and highlighting are used to enhance groupings of similar rows or columns in a table (Tullis 1988). However, effective spacing of rows or columns in a table can enhance information extraction tasks without the use of color. Therefore, color displays should only facilitate extraction of information from complex line graphs and bar graphs when discrimination and perceptual grouping are requisite components of the task.

2.2 Experimental Design

An empirical study tested these general hypotheses regarding the appropriateness of a particular graphic for a specific task and the effect of color. The study used a within-subjects factorial design in which each subject was measured under all factorial combinations of four independent variables: display format, color, task, and data set. The study used three levels of **display format**: bar graph, line graph and table; two levels of **color**: color or monochrome (white on black background); and three levels of **task**: point reading, comparison, and trend questions (Tan and Benbasat 1990). The study also used five different **data sets**. The data sets contained selected data from *Statistical Abstract of the United States* (Census Bureau 1990). All the data sets expressed six categories of information over twelve time periods. Thus, each subject answered three questions, one per task, from thirty unique displays (3 display formats x 2 color x 5 data sets), yielding a total of ninety observations per subject.

Decision accuracy and decision time, measured in milliseconds (msec), are the dependent variables. Eye movement process data was also collected to provide a componential analysis of decision time. Each eye fixation was categorized into one of six categories: legend, x-axis, y-axis, data values, frame or question. Thus, total decision time is the sum of the time spent examining the legend, using the x-axis, looking at data values, etc. These classes are adopted from Kosslyn (1989). For bar and line graphs, the legend associates a pattern, color, or symbol to the categorical name. For tables, the legend is the categorical name. The x-axis presented labels for a time series (either months or years). The y-axis labeled the scaled values of the data series. In tables, the data values are the numeric entries indexed by row and column labels in a table. In bar and line graphs, the data values are encoded symbolically using colored or textured bars and colored line segments with symbols. The frame included a box around the data values. For bar and line graphs, the frame also included the tick marks used to locate labels on the x and y axes. The category, "question" was the yes or no question posed to the display.

2.3 Collecting Process Tracing Data Using Eye Movements

The Eyegaze System from LC Technologies (Fairfax, Virginia) captured eye movement data. The Eyegaze System uses an Intel-80486 based personal computer to operate both the eye movement tracing system and the application software simultaneously. The Eyegaze System uses the pupil-center/corneal reflection method to determine eye gaze (Young and Sheena 1975). A video camera, sensitive in the infra-red range and positioned below the computer monitor, continually observes the subject's eye (Figure 1). A small, low power, infrared, light emitting diode (LED) located at the center of the video camera lens illuminates the user's eye. The LED generates corneal reflection and causes the bright pupil effect which enhances the camera's image of the pupil. By means of video image processing, an algorithm determines the center of the pupil and the brightest reflection of the cornea (as illuminated by the LED). An algorithm computes the distance between these two points. This distance is related linearly to changes in the observer's gaze point and is independent of small movements of the head providing the eye remains in the video camera field of view. Trigonometric calculations determine the subject's gaze point based on the positions of the pupil center and the corneal reflection within the video image. Specialized image-processing software generates x, y coordinates for the intercept of the gaze line on the monitor screen as well as other measures, including fixation duration, pupil diameter, and eye blinks. The observer's eye is about 20 inches from the screen of the computer monitor. No attachments to the head are required.

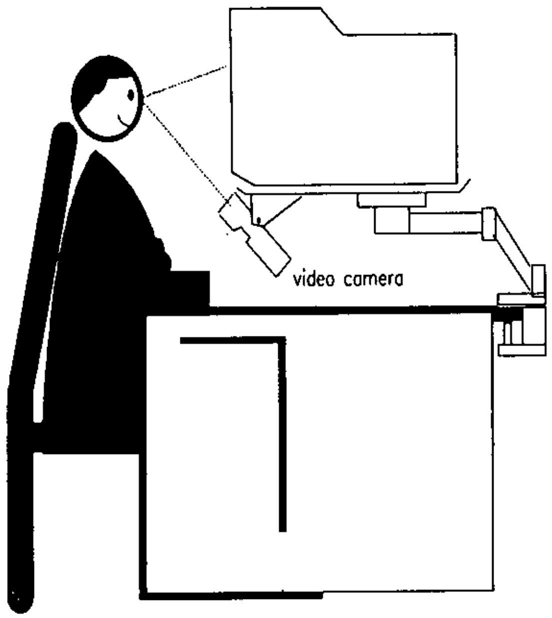

Figure 1. The Eyegaze System from LC Technologies for Tracing and Recording Eye Movements

2.4 Procedure

Each subject participated in three sixty minute sessions. Each session was twenty-four hours apart. At the first session, subjects were screened to determine whether the Eyegaze System could track their gaze. Excluded subjects were color blind, wore bifocals or trifocals, had a "droopy" eyelid, or were sensitive to the LED light source. About 90% of the subjects recruited to participate in the study could be calibrated. Once calibrated, subjects played video games using their eye to control objects in the game. This practice helped subjects learn to make eye movements with minimal head movement.

After fifteen minutes of video games, subjects completed a practice exercise containing fifteen yes/no questions posed to graphic displays. Data were collected using a computer program for MS-DOS compatible PC microcomputers that posed a question and a display to the subject. Subjects read and studied each question *without* the time pressure of the stopwatch, although question reading time was measured. Presenting the question before the graphic display was shown allowed subjects to extract the important semantic cues needed to answer the question and to retain this information in working memory. However, the question also was displayed on the second screen that presented the graph or table. Subjects answered each question based on information presented on a graph or table. The program collected decision time for each question. Each question only required a few seconds (averaging three to seven seconds) to answer. If subjects answered the question incorrectly, either in the practice session or in the subsequent sessions, the system generated a short beep as feedback before continuing to the next question. Figure 2 illustrates one subject's sequence of eye movements used to answer a comparison question posed to a bar graph.

At the second session, subjects first completed a practice session and then answered forty-five yes/no questions representing the fully crossed factorial combinations of three types of tasks (read, compare, trend), three types of displays (bar graph, line graph, table) and five different data sets. Half of the displays were monochrome; the other half were color. Between each question for this study, subjects answered a question regarding a different study being conducted concurrently. Subjects answered ninety questions across both studies. Thus, each subject answered a total of 180 questions during the second and third sessions. Alternating the questions helped reduce memorization effects. Displays were randomly ordered with the stipulation that a question regarding the same task from the same data set be at least twenty questions apart.

At the third session, subjects once again completed a practice session containing fifteen questions and then answered forty-five yes/no questions posed to the various task x graph x data set combinations. If the display for a particular task x graph x data set combination was monochrome during the second session, it became a color display for the third session and visa versa. Further, if the correct answer to a question was "yes" during the second session, the question was reworded to make the correct answer become "no" for the third session. This was done to reduce potential memorization effects. Question order for all sessions was completely randomized for each subject.

2.5 Subjects

Twenty undergraduate business students who were not color-blind participated in the study. Participation was voluntary. Subjects earned a base rate of $9.00 in addition to incentives based on performance. Subjects could earn up to $0.10 per question and they lost $0.25 for each question answered incorrectly. For each question, subjects earned the dollar equivalent of (20-RT)/200. RT represents the reaction time in seconds. If a subject required ten seconds to answer a question, earnings for that question would be $0.05. Earnings ranged from $16 to $24 dollars (over both studies) and averaged $20 for approximately three hours of participation.

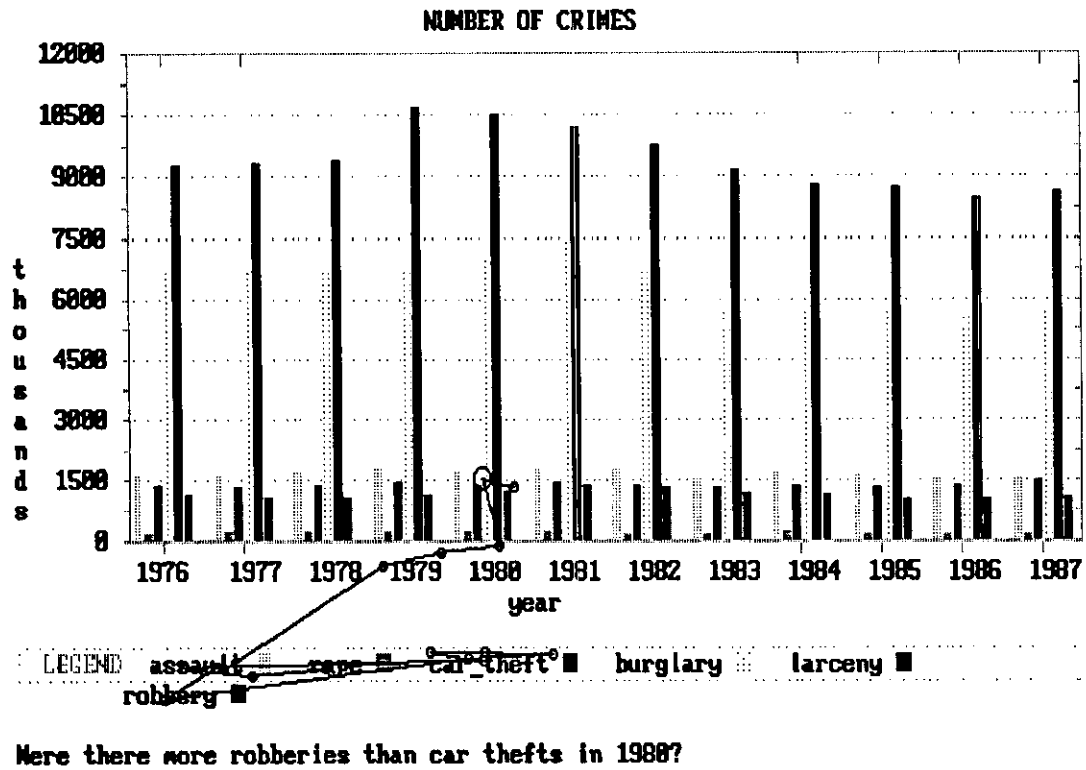

Figure 2. A Subject's Sequence of Eye Movements Used to Answer a Comparison Question Posed to a Bar Graph

3. DATA ANALYSIS AND RESULTS

Twenty subjects each answered questions posed to factorial combinations of graph format, color, task, and data set resulting in 1,800 observations (20 x 3 x 2 x 3 x 5). The first set of analyses examined interaction effects among graph, color, and task for the dependent variables accuracy and decision speed (seconds). The second set of analyses examined over 35,000 eye fixations. This represents approximately twenty fixations per observation in the first set of analyses. These more detailed eye movement data provide information regarding where subjects looked to answer the questions. Specifically, the fixation data were aggregated to provide the total time spent in six regions of the display: legend, x-axis, y-axis, data values, frame or question.

Effects were tested using the following analysis of variance model for factorial within-subjects designs. Each effect to be tested has its own error term. The denominator for each F test is the interaction term with subject, π (see Howell 1987, pp 452-456).

$$Y_{ijklmn} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \phi_m + \pi_n + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk}$$
$$+ \alpha\beta\pi_{ijn} + \alpha\gamma\pi_{ikn} + \beta\gamma\pi_{jkn} + \alpha\beta\gamma_{ijk} + \alpha\beta\gamma\pi_{ijkn} + \varepsilon_{ijklmn}$$

where:

μ is constant,

α_i is the effect of graph format (i = 1,...3),

β_j is the effect of task (j = 1,...3),

γ_k is the effect of color (k = 1,2),

δ_l is the effect of data set (l = 1,...5),

ϕ_m is the effect of order (m = 1,2),

π_n is the experimental participant (n = 1,...20), and

ε_{ijklmn} is the experimental error term.

3.1 Accuracy

Overall, the mean error rate was 4.67%. The ANOVA found a significant three-way interaction between graph, task and color ($F(4,76) = 5.19$), $p<.0004$) and a significant two-way interaction between graph and task ($F(4,76) = 7.21$), $p<.0001$). No other two-way interaction effects were significant. The ANOVA did not find any significant main effects for graph, task, color, order or data set. Inspection of the treatment means found that the error rate using monochrome bar graphs for answering trend questions (20%) was significantly higher than the error rate using color bar graphs as well as for most other treatments (Figure 3). For answering point reading questions, monochrome line graphs had a significantly higher error rate than some other treatments. Inspection of Figure 3 does not reveal any other insightful patterns for the three-way interaction effects.

This error rate is slightly higher than the typical error rate of 3% in response to normal instructions for studies in experimental psychology (Pachella 1984). Thus, the analyses on the original set of 1,800 observations examined the data for possible speed-accuracy trade-offs. Overall, the correlation between reaction time for each decision and the probability of an error was 0.092. This positive correlation is significantly greater than zero ($p<.0001$). In no case was there a significant negative correlation between reaction time and error. Negative correlations demonstrate that a speed-accuracy tradeoff exists. Thus, these data are free from any concerns with speed-accuracy tradeoffs and only error-free data were used in the subsequent data analysis as this is the standard practice for cognitive modeling studies (e.g., Card, Moran, and Newell 83).

3.2 Decision Speed

There was not a significant three-way interaction between graph, task and color ($F(4,76) = 1.94$, $p>.10$). The ANOVA did not find a significant interaction between graph and task ($F(4,76) = 2.29$, $p>.0849$). Decision speed was not significantly different among trend or comparison questions answered using bar or line graphs. However, for all tasks (point reading, comparison, and trend), tables were faster than graphs ($F(2,38) = 69.81$, $p<.0001$). For all graphs (bar, line and table), point reading tasks were the fastest ($F(2,38) = 58.56$, $p<.0001$). Figure 4a shows the graph x task treatment means.

The ANOVA also found a significant interaction between graph and color ($F(2,38) = 20.16$, $p<.0001$) and task and color ($F(2,38) = 6.01$, $p<.0047$). Figure 4b shows the graph x color treatment means. Compared to monochrome displays, color significantly increased the decision speed for bar graphs and line graphs; however, color did not increase the decision speed of questions posed to tabular displays. Figure 4c shows the task x color treatment means. Color only increased the decision speed for trend questions; color did not increase decision speed of point reading or comparison questions.

3.3 Componential Analyses of Decision Speed

Eye fixation data enable a separate analysis of the time spent examining six parts of the graph or table. These include the legend, x-axis, y-axis, data values, frame and question. The ANOVA conducted for time spent looking at the legend found a significant interaction between graph and task ($F(4,76) = 31.20$, $p<.0001$). Tables average only 430 msec; bar and line graphs averaged 1,735 msec (Figure 5). The ANOVA conducted for the time spent looking at the data values found a significant three-way interaction between graph, task, and color ($F(4,76) = 4.07$, $p<.0027$). Examination of the treatment means found no difference between color or monochrome tables for any table, but a significant difference between color or monochrome bar and line graphs for comparison and trend tasks.

3.4 Predicting Variance in Reaction Times

The ANOVA conducted for accuracy ($R^2=.2321$) and decision speed ($R^2=.4978$) explained a significant portion of the variance. Unfortunately, the models are not very parsimonious; 459 degrees of freedom are attributed to the model. The eye fixation data provide a more parsimonious approach. A simple linear regression was used to predict reaction time in seconds given *the actual number of fixations* a subject used to answer the question using the model:

$$Y_i = \mu + \alpha_i + \varepsilon$$

where:

μ is constant, and

α_i is the actual number of fixations per question.

The regression explained over 85% of the variation in reaction time using a single parameter. Moreover, the addition of the experimental factors (graph, task, color, subject, data set, order) and their interactions did not significantly increase the R-square ($R^2_{reduced}=.8543$ to $R^2_{full}=.8582$). In the study, subjects pressed the space bar or return key to indicate a "yes" or "no" response to the question. The parameter estimate for the intercept was 330 msec. This value agrees with published values for time required to enter a keystroke (Card, Moran, and Newell 1983; Olson and Olson 1990). The parameter estimate, 269 msec, agrees with the published values for the eye movement and dwell time associated with each fixation (Card,

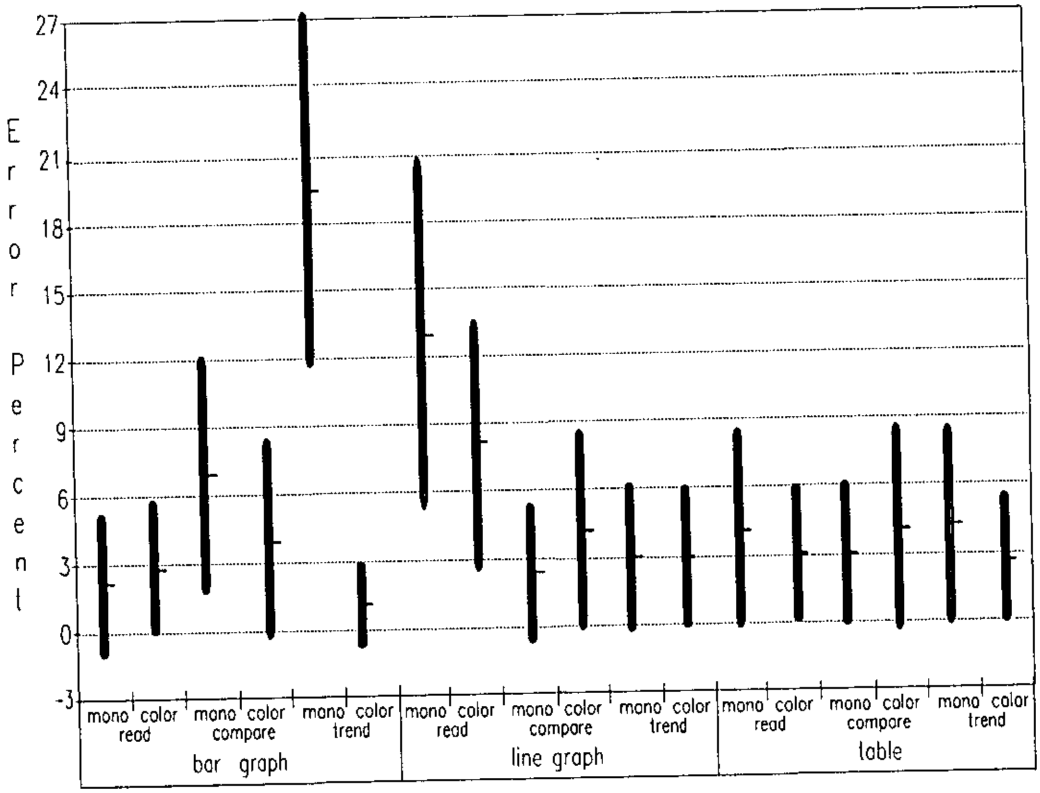

Figure 3. Mean Accuracy of Three-Way Interactions for Graph, Task, and Color with 95% Confidence Intervals

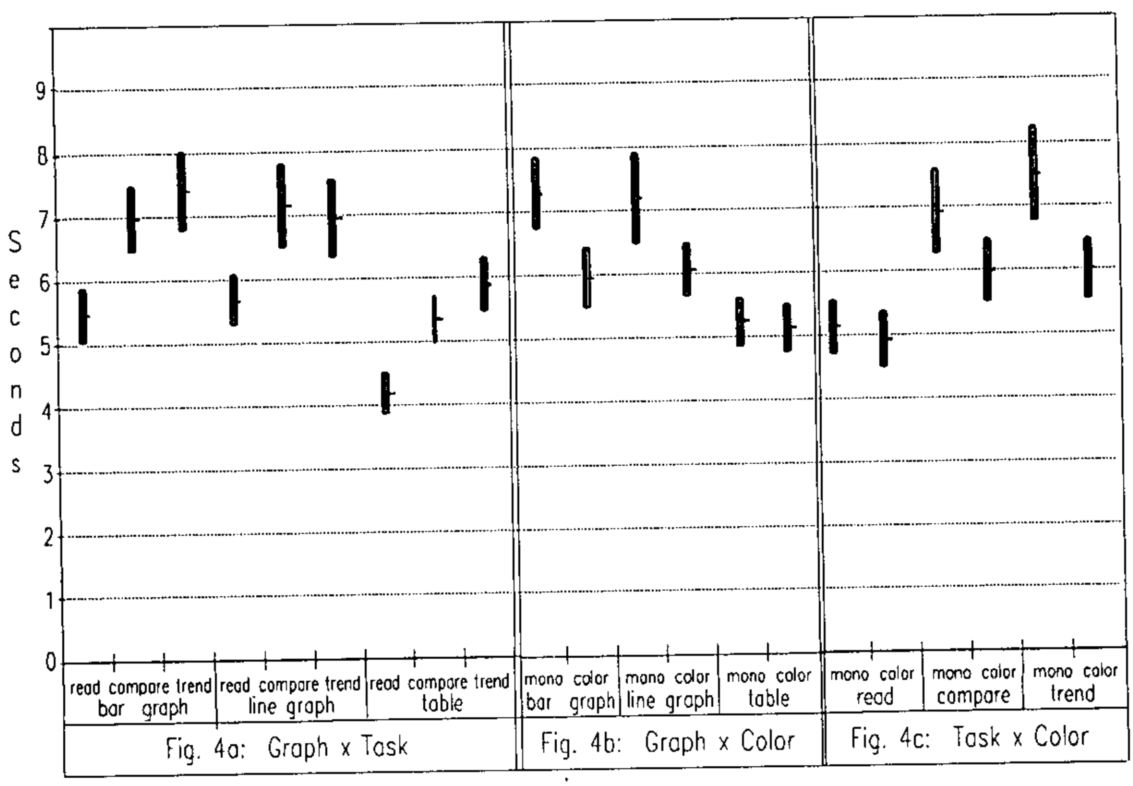

Figure 4. Decision Speed Means of Two-Way Interactions for Graph, Task, and Color with 95% Confidence Intervals

Table 1. Comparison of Mean Time Spent Examining Eye Fixations on Data Values (Seconds). Excludes the time spent examining the legend, x-axis, y-axis, etc.

DISPLAY	TASK	COLOR	N	Mean	SE
bar graph	read	B/W	97	2.00	0.08
		color	97	1.79	0.09
	compare	B/W	93	3.00	0.14
		color	96	2.17	0.10
	trend	B/W	79	4.57	0.24
		color	99	2.96	0.11
line graph	read	B/W	87	2.25	0.13
		color	92	2.01	0.12
	compare	B/W	98	2.98	0.21
		color	96	2.26	0.11
	trend	B/W	97	3.85	0.26
		color	97	2.69	0.10
table	read	B/W	96	2.02	0.10
		color	97	1.97	0.10
	compare	B/W	97	2.76	0.13
		color	96	2.72	0.12
	trend	B/W	96	2.82	0.12
		color	97	2.73	0.13

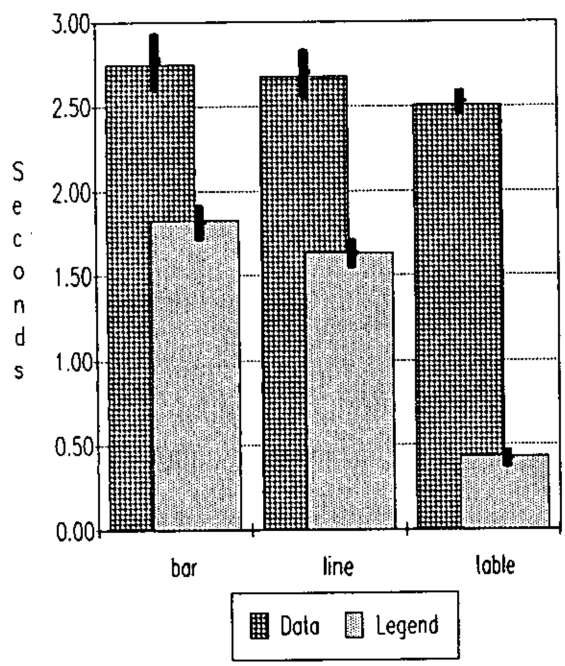

Figure 5. Mean Time Spent Examining Data Values and the Legend with 95% Confidence Intervals

Moran, and Newell 1983; Olson and Olson 1990; Russo 1978). Of course, using *the actual number of eye fixations* to predict reaction time overstates the explanatory power of this approach. However, this simple regression provides insight about modeling reaction time to answer a question posed to a graphic display as a function of *the predicted number of fixations*, where: reaction time = 330 msec + 269 msec * number of fixations. This quantitative approach is discussed in the next section.

4. DISCUSSION

If the prevailing guidelines for selecting graphic displays for a particular task are correct, tables should be faster and more accurate for reading specific data values, bar graphs should be faster and more accurate for comparison tasks, and line graphs should be faster and more accurate for evaluating trends. These recommendations by Jarvenpaa and Dickson and predictions by Vessey conflict with the results reported in this paper. The current study found one major difference in performance accuracy: the error rate from using monochrome bar graphs for answering trend questions was significantly higher than the error rate from using other line graphs or tables. For this single case, contingency theory or the theory of cognitive fit seems to apply. There was not congruence between the task and the decision aid, hence performance accuracy decreased. However, in all other evaluations of accuracy, differences among bar graphs, line graphs, and tables were not contingent upon the task. Furthermore, the analyses of decision speed found that tables were always faster than graphs regardless of the task.

Interpretation of the eye fixation results helps explain this phenomenon. Tables are comprised of alphanumeric characters. No legends are needed to map symbols, colors, or textures to categorical labels; the legend is the column label. The row labels are equivalent to the x-axis labels on the bar and line graphs. The row and column labels index all of the entries in the table. All comparisons are made at the semantic level. The elementary graphical perception tasks made directly from comparing heights of bars or slopes of lines were as fast from graphs as the comparison of two numbers in a table (Figure 5). However, the time required to associate the symbol, color, or texture in the legend to the categorical name significantly increased the time spent evaluating the legend using bar and line graphs in comparison to tables. Thus, the total time for bar and line graphs was slower than that for tables.

It is also important to note that differences between color and monochrome displays were found only between color and monochrome bar or line graphs for comparison and trend tasks. No differences were found for tables; no differences were found for point reading tasks. Thus, color facilitated discrimination tasks about the data values within a display, but did not affect performance when difficult discrimination tasks were not required.

Needless to say, these results do not suggest that tables are always better than graphs. Any advantage of tables over graphics is still contingent upon the task. Had the task required subjects to make comparisons between trends, the large cognitive overhead necessary to process this information from tables may have enabled graphs to perform better than tables. Further, other graphic display designs (e.g., with a different scale or with legend labels directly on each line or bar) may reduce the difficulty of the graphic information processing for the tasks used in the current study.

This raises an important issue. Seemingly subtle graphic design changes or changes in the complexity of the task affect decision speed and decision accuracy. These effects are pervasive throughout the behavioral decision making literature. For example, changes in presentation format between decimals and fractions induce preference reversals (Johnson, Payne and Bettman 1988), graphic versus tabular formats alter framing effects (Diamond and Lerch 1992), and changes in graph scale influence decision performance (Taylor and Anderson 1986) These examples demonstrate that subtle changes in information format influence information processing behavior as well as subsequent decision making. In fact, Jarvenpaa (1989, p. 300) states that "the main payoffs for decision aid research are likely to come from research that articulates ***how*** particular features of the aids enhance human strengths or remedies (sic) human weaknesses in extracting and using information." Thus, it is important to predict how any change in graphic design or task complexity would affect performance.

Although qualitative theories provide a general basis for making predictions, they are difficult to apply in these more subtle contexts because one can not ascertain whether an appropriate graph has been selected for a particular task. There is no metric to quantify the fit between the graphic decision aid and the task nor a metric to measure the quality of a specific graphic design objectively. Further, these theories make no predictions regarding the use of color displays. Thus, qualitative theories enable one to argue that any evaluation could be biased because the graphics were designed poorly or that the task was too complex or too simple. Such caveats suggest that the next generation of MIS graphics research adopt a more objective approach for predicting the efficacy of a graphic design for a specific task.

The eye fixation data show that the specific information processing requirements of the task vary as a function of the specific nature of the display. By determining fixation by fixation information processing requirements, one can quantify performance. The simple linear regression model from section three used a single parameter, actual number of eye fixations, and explained over 85% of the variation in reaction times compared to less than 50% for traditional models. Furthermore, fixation data account for individual differences between subjects as well as potential interactions among graph, task, and color.

A predictable sequence of eye fixations provide a detailed account of the number of information processing subtasks used to acquire information from the display. The complexity and number of these subtasks are a function of the task and the graph format. Computational models of cognition have defined a timing parameter for most of these subtasks (Card, Moran, and Newell 1983). Decision time can be estimated from a summation of the timing parameters assigned to each subtask. This approach quantifies how subtle changes in the graph format or the task can influence the cognitive information processing burden.

5. SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

The graphs versus tables controversy focused research attention on empirically evaluating which display tool is best for particular tasks. While the ensuing research has provided numerous insights, it is neither possible nor desirable to resolve each and every graphic display design issue with an empirical study. The goal of such research has been to enhance our understanding of how people process information from displays. Presumably, this knowledge would help the MIS community design better graphic decision aids. Unfortunately, general qualitative predictions and guidelines are difficult to apply in the context of designing a specific graphic for a given task. Graphics research needs a more objective theory of graphics that provides robust quantitative predictions for evaluating alternative graphic designs.

Very little MIS graphics research could be codified into an algorithm to provide robust quantitative predictions about the effectiveness of a graphic decision aid for a specific task. Keen (1980) cautioned others against using qualitative contingency theories because they are untestable; the results are always subject to caveats regarding potential interactions among other factors in the study. This does not deny the value of contingency theories. Keen only noted that they are hard to challenge and difficult to apply to real problems. In Newell's view (Card, Moran, and Robertson 1992), if psychological theories are to be usable, the mechanisms underlying performance should be teased out using computational models of behavior. Computational cognitive models should emphasize detailed task analysis, quantitative calculation, and zero-parameter predictions to make them useful for real applications. GOMS models are one example of such an approach (Card, Moran, and Newell 1983).

This paper motivates a computational cognitive engineering approach for quantifying the effectiveness, efficiency and quality of graphic designs. Unlike prior MIS graphics research, a quantitative cognitive modeling approach can provide robust objective predictions using an algorithm that could be incorporated into software for automated graphic design. Quantitative models provide a falsifiable theory of graphics that can be empirically tested by comparing decision time predictions from the model with actual decision times. Thus, the next generation of graphics research not only should enhance our understanding of graphical perception but also provide quantitative predictions that facilitate the objective evaluation of the effectiveness, efficiency and quality of graphic displays.

6. ACKNOWLEDGMENTS

This material is based upon work supported by the Information, Robotics, and Intelligent Systems Division of the National Science Foundation under Grant No. IRI-9209576 and Grant No. 3-71857 from the University of Pennsylvania Research Foundation. I am grateful to five anonymous reviewers for their helpful comments and suggestions that helped me reorganize an earlier draft of this paper.

7. BIBLIOGRAPHY

Benbasat, I.; Dexter, A. S.; and Todd, P. "The Influence of Color and Graphical Information Presentation in a Managerial Decision Simulation." *Human-Computer Interaction*, Volume 2, 1986, pp. 65-92.

Bertin, J. *Sémiologie graphiques*, Second Edition. Paris, France: Gauthier-Villars, 1967. [English translation: W. J. Berg, *Semiology of Graphics*. Madison, Wisconsin: University of Wisconsin Press, 1983.]

Card, S. K.; Moran, T. P.; and Newell, A. *The Psychology of Human-Computer Interaction*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., 1983.

Card, S. K.; Moran, T. P.; and Robertson, G. "Remembering Allen Newell." *SIGCHI Bulletin*, Volume 24, Number 4, 1992, pp. 22-24.

Casner, S., and Larkin, J. H. "Cognitive Efficiency Considerations for Good Graphic Design." In *Eleventh Annual Conference of the Cognitive Science Society*, University of Michigan, Ann Arbor, Michigan. Hillsdale, New Jersey: Lawrence Erlbaum Associates., 1989, pp. 275-282.

Census Bureau. *Statistical Abstract of the United States*, 110th Edition. Washington, DC: USA Bureau of the Census, 1990.

Cleveland, W. S. *The Elements of Graphing Data*. Monterey, California: Wadsworth Publishing, 1985.

DeSanctis, G. "Computer Graphics as Decision Aids: Directions for Research." *Decision Sciences*, Volume 5, Number 3, 1984, pp. 463-487.

Diamond, L., and Lerch, F. J. "Fading Frames: Data Presentation and Framing Effects." *Decision Sciences*, Volume 23, Number 5, 1992, pp. 1050-1071.

Howell, D. C. *Statistical Methods for Psychology*, Second Edition. Boston: Duxbury Press, 1987.

Ives, B. "Graphical User Interfaces for Business Information Systems." *MIS Quarterly*, Special Issue, 1982, pp. 15-47.

Jarvenpaa, S. L. "The Effect of Task Demands and Graphic Format on Information Processing Strategies." *Management Science*, Volume 35, Number 3, 1989, pp. 285-303.

Jarvenpaa, S. L., and Dickson, G. W. "Graphics and Managerial Decision Making: Research Based Guidelines." *Communications of the ACM*, Volume 31, Number 6, 1988, pp. 764-774.

Johnson, E. J.; Payne, J. W.; and Bettman, J. R. "Information Displays and Preference Reversals." *Organizational Behavior and Human Decision Processes*, Volume 42, 1988, pp. 1-21.

Keen, P. G. W. "MIS Research: Reference Disciplines and a Cumulative Tradition." In E. McLean (Ed.), *Proceedings of the First International Conference on Information Systems*, Philadelphia, Pennsylvania, 1980, pp 9-18.

Kinney, J. S., and Huey, B. M. *Application Principles for Multicolor Displays.* Washington, DC: National Academy Press, 1990.

Kleinmuntz, D. N., and Schkade, D. A. "The Cognitive Implications of Information Displays in Computer-Supported Decision Making." Working paper #2010-88, Sloan School of Management, Massachusetts Institute of Technology, 1989.

Kosslyn, S. M. "Understanding Charts and Graphs." *Applied Cognitive Psychology*, Volume 3, 1989, pp. 185-225.

Legge, G. E.; Gu, Y.; and Luebker, A. "Efficiency of Graphical Perception." *Perception and Psychophysics*, Volume 46, 1989, pp. 365-374.

Lohse, G. L. "A Cognitive Model for the Perception and Understanding of Graphs." In S. P. Robertson, G. M. Olson, and J. S. Olson (Eds.), *CHI'91 Conference Proceedings*, New Orleans, Louisiana. Los Alamitos, California: ACM Press, 1991a, pp. 137-144.

Lohse, G. L. "A Cognitive Model for Understanding Graphical Perception." Cognitive Science and Machine Intelligence Laboratory Technical Report, 39, The University of Michigan, Ann Arbor, 1991b.

Mackinlay, J. D. "Automating the Design of Graphical Presentations of Relational Information." *ACM Transactions on Graphics*, Volume 5, Number 2, 1987, pp. 110-141.

Olson, J. R., and Olson, G. M. "The Growth of Cognitive Modeling in Human-Computer Interaction Since GOMS." *Human-Computer Interaction*, Volume 5, 1990, pp. 221-265.

Pachella, R. G. "The Interpretation of Reaction Time in Information Processing Research." In B. Kantowitz (Ed.), *Human Information Processing: Tutorials in Performance and Cognition*. New York: Halstead Press, 1984, pp. 41-82.

Pinker, S. "A Theory of Graph Comprehension." In R. Freedle (Ed.), *Artificial Intelligence and the Future of Testing*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1990, pp. 73-126.

Russo, J. E. "Adaptation of Cognitive Processes to Eye Movement Systems." In J. W. Senders, D. F. Fisher, and R. A. Monty (Eds.), *Eye Movements and Higher Psychological Functions*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1978, pp. 89-109.

Spence, I. "Visual Psychophysics of Simple Graphical Elements." *Journal of Experimental Psychology: Human Perception and Performance*, Volume 16, Number 4, 1990, pp. 683-692.

Spence, I., and Lewandowsky, S. "Displaying Proportions and Percentages." *Applied Cognitive Psychology*, Volume 5, 1991, pp. 61-77.

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

Tullis, T. S. *Display Analysis Program, Version 4.0* [Computer Program]. Available from The Report Store, Lawrence, Kansas, 1986.

Tullis, T. S. "Screen Design." In M. Helander (Ed.), *Handbook of Human-Computer Interaction*. New York: Elsevier Science Publishers, Inc., 1988.

Tan, J. K. H, and Benbasat, I. "Processing of Graphical Information: A Decomposition Taxonomy to Match Data Extraction Tasks and Graphical Representations." *Information Systems Research*, Volume 1, Number 4, 1990.

Tan, J. K. H, and Benbasat, I. "Understanding the Effectiveness of Graphical Presentation for Information Extraction: A Cumulative Experimental Approach." *Decision Sciences*, in press.

Taylor, B. G., and Anderson, L. K. "Misleading Graphs: Guidelines for the Accountant." *Journal of Accountancy*, October 1986, 126-135.

Vessey, I. "Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature." *Decision Sciences*, Volume 22, Number 2, 1991, pp. 219-240.

Young, L. R., and Sheena, D. "Survey of Eye Movement Recording Methods." *Behavior Research Methods and Instrumentation*, Volume 7, Number 5, 1975, pp. 397-429.