User Model User-Adap Inter (2006) 16: 1–30
DOI 10.1007/s11257-006-9002-9

ORIGINAL PAPER

A model of perceptual task effort for bar charts and its role in recognizing intention

Stephanie Elzer · Nancy Green · Sandra Carberry · James Hoffman

Received: 1 June 2004 / Accepted in revised form:
4 January 2006 / Published online: 30 May 2006

Abstract This paper presents a model of perceptual task effort for use in hypothesizing the message that a bar chart is intended to convey. It presents our rules, based on research by cognitive psychologists, for estimating perceptual task effort, and discusses the results of an eye tracking experiment that demonstrates the validity of our model. These rules comprise a model that captures the relative difficulty that a viewer would have performing one perceptual task versus another on a specific bar chart. The paper outlines the role of our model of relative perceptual task effort in recognizing the intended message of an information graphic. Potential applications of this work include using this message to provide (1) a more complete representation of the content of the document to be used for searching and indexing in digital libraries, and (2) alternative access to the information graphic for visually impaired users or users of low-bandwidth environments.

Keywords Perceptual effort · Cognitive modeling · Diagrams · Plan recognition

S. Elzer (✉)
Dept of Computer Science, Millersville Univ., Millersville, PA 17551 USA
e-mail: elzer@cs.millersville.edu

N. Green
Dept of Mathematical Sciences, University of North Carolina at Greensboro, Greensboro, NC 27402, USA
e-mail: nlgreen@uncg.edu

S. Carberry
Department of Computer Science, University of Delaware, Newark, DE 19716 USA
e-mail: carberry@cis.udel.edu

J. Hoffman
Department of Psychology, University of Delaware, Newark, DE 19716 USA
e-mail: hoffman@udel.edu

1 Introduction

Information graphics (line graphs, bar charts, etc.) are pervasive in popular media such as newspaper and magazine articles. Such graphics enable complex information to be assimilated perceptually with ease. As more and more of the documents containing information graphics become available electronically, the challenge is to develop techniques for providing effective access so that the information is readily available when needed, and so that all individuals can benefit from these resources. For example, information graphics pose a problem when attempting to search the content of mixed-media publications within digital libraries. The searchable index of such documents should include not only the important content of the text, but also of the information graphics contained in the document. Such graphical information is also not readily accessible in environments with low-bandwidth transmission or miniature viewing facilities. For example, recent developments have led to wireless telephones for accessing the Web. However, the low bandwidths and small screens of these devices limit effective retrieval of graphical information. Furthermore, individuals with impaired eye-sight have limited access to graphics, thus preventing them from fully utilizing available information resources. In order to provide effective access to the information contained in information graphics, there is a need for tools that convey the content of information graphics in other modalities.

Although some information graphics are only intended to display data values, our analysis of a corpus of information graphics from popular media sources indicates that such information graphics generally have a communicative goal and that this intended message is often not conveyed by accompanying text. For example, the graphic in Fig. 1 ostensibly is intended to convey that Delaware personal bankruptcies rose in 2001 in contrast with the preceding trend from 1998 to 2000. Moreover, the design choices made by the graphic designer can impact what a graphic conveys. For example, drug spending data for different age groups could be displayed so that it is sorted by age, thereby eliciting any trends in drug spending over the average person's lifetime (see Fig. 2a). The same data could also be sorted by the amount spent, thereby eliciting the observation of the age ranges with the maximum and minimum spending (see Fig. 2b). Although this design looks unusual, graphics do occur in which the bars have numeric

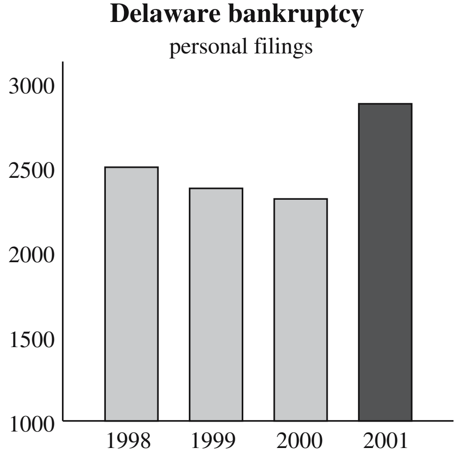
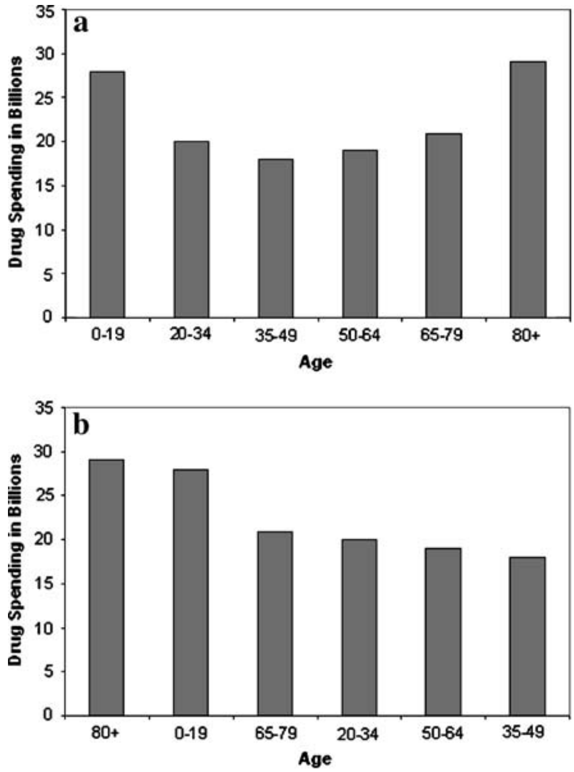


Fig. 1 Graphic from the Wilmington News Journal (a local newspaper)

Fig. 2 Two alternative graphs from the same data[1]



labels but are ordered by bar height. For example, our corpus includes a graphic from USA Today in which the bars are labeled with the release dates of movies but the bars appear in the graphic in order of revenue, which is depicted by their height. As noted by Iverson and Gergen (1997) and Green et al. (2004), a set of data can be presented in many different ways, and graphs are often used as a communication medium or rhetorical device for presenting a particular analysis of the data and enabling the viewer to better understand this analysis. Thus recognizing the intended message of an information graphic is crucial for full comprehension of a mixed-media resource.

One might expect that a graphic's caption would provide its intended message. However, Corio and Lapalme (1999) performed a large corpus study of information graphics and noted that captions often do not give any indication of what the information graphic conveys. Our examination of a collection of graphics supports their findings. Thus we must be able to infer the message underlying a graphic when captions are missing or of little use.

We hypothesize that the core message of an information graphic (the primary overall message that the graphic conveys) can serve as the basis for an effective summary. This summary could then be used in a variety of ways. For digital libraries, the summary of the information graphic could be used to appropriately index the graphic and to enable intelligent retrieval. If there is accompanying text, the summary of the

[1] The data displayed in this graphic is for illustration purposes only and is not based on factual findings.

graphic can be used in conjunction with a summary of the document's text to provide a more complete representation of the document's content. For individuals who are sight-impaired or who are using low-bandwidth devices, using the core message of the information graphic as the basis for a summary would provide access to the informational content of the graphic in an alternative modality. Rather than providing alternative access to what the graphic looks like or a listing of all of the data points contained in the graphic, our approach would provide the user with the message and knowledge that one would gain from viewing the graphic.

The problem of graph summarization has received some attention, although in different contexts than our work. Yu et al. (2002) have used pattern recognition techniques to summarize interesting features of time-series data from a gas turbine engine. In a related project, Sripada et al. (2002) have looked at time-series data in the domain of weather forecasting. The purpose of the graphs in these domains[2] is to present information that domain experts can analyze and use—the fact that there is no communicative intention realized in this systematic presentation of raw data distinguishes it from the information graphics that we are studying. Futrelle and Nikolakis (1995) developed a constraint grammar formalism for parsing vector-based visual displays and producing structured representations of the elements comprising the display. The goal of Futrelle's project is to produce a graphic that summarizes one or more graphics from a document (Futrelle 1999). The summary graphic might be a simplification of a graphic or a merger of several graphics from the document, along with an appropriate summary caption. Thus the end result of summarization will itself be a graphic.

This paper first outlines our overall approach to inferring the communicative message of an information graphic and briefly describes the various types of evidence that can aid the inference process. It then focuses on one specific type of evidence, perceptual task effort, and the work that we have done in order to rank perceptual tasks in terms of the effort required on a particular bar chart. It describes our rules for estimating perceptual task effort, and presents the results of an eye tracking experiment conducted in order to evaluate the rankings provided by our effort estimates. We conclude with an example that illustrates the impact of perceptual task effort on our system for inferring the communicative message of an information graphic. Our current work focuses on simple bar charts. By simple bar charts, we mean bar charts that display the values of a single independent attribute and the corresponding values for a single dependent attribute. In future work, we will expand our model to encompass other graph types, such as line graphs, pie charts, and grouped or stacked bar charts.

2 Recognizing the graphic designer's intended message

As Clark (1996) noted, language is more than just words. It is any "signal" (or lack of signal when one is expected), where a signal is a deliberate action that is intended to convey a message. Language research has posited that a speaker or writer executes a speech act whose intended meaning he expects the listener to be able to deduce, and that the listener identifies the intended meaning by reasoning about the

[2] It should be noted that processing did not start from the graphic but rather from the data tables underlying the graphic (Sripada et al. 2002; Yu et al. 2002).

observed signals and the mutual beliefs of author and interpreter (Grice 1969; Clark 1996). Applying Clark's view of language to information graphics, it is reasonable to presume that the author of an information graphic similarly expects the viewer to deduce from the graphic the message that he intended to convey by reasoning about the graphic itself, the salience of entities in the graphic, and mutual beliefs. Thus while the AutoBrief project extended speech act theory to the *generation* of information graphics (Kerpedjiev and Roth 2000; Green et al. 2004), we propose to extend it to the *understanding* of information graphics.

Beginning with the seminal work of Perrault and Allen (1980) who developed a system for deducing the intended meaning of an indirect speech act, researchers have applied plan inference techniques to a variety of problems associated with understanding utterances, particularly utterances that are part of a dialog. In our research, we have extended plan inference techniques that have been used successfully on natural language discourse to inferring intention from information graphics (Elzer et al. 2005). Following the work of Charniak (Charniak and Goldman 1993) and others, we capture plan inference in a Bayesian belief network. A Bayesian belief network is a probabilistic framework based on Bayes' rule, and it is also sometimes referred to as a belief network, a Bayesian network, or a causal network. Without considering the domain or intended application of the network, a Bayesian network can be viewed as consisting of a graphical structure which encodes a domain's variables and the *qualitative* relationships (often causal) between those variables as well as the *quantitative* probabilities over the variables (Pearl 1988; Druzdzel and van der Gaag 2000). In a Bayesian belief network used for plan inference, the nodes in the network represent propositions and the arcs between the nodes represent causal dependencies. These dependencies are captured using conditional probability distributions that represent the likelihood of a proposition given the various values of its parent node(s). Bayes' rule is used to compute the probability of each proposition given causal evidence (from its parents) and diagnostic evidence (from its children). The root nodes of the network represent various high-level hypotheses about an agent's plan (Carberry 2001).

In all plan inference systems, there is some explicit plan structure, which defines the relationships among goals, subgoals, and primitive actions. In Bayesian networks, the plan structure is captured by the network itself; each goal, subgoal, and primitive action is represented as a piece of a network. If a goal can be decomposed into a particular set of subgoals or primitive actions, an arc from the goal to each subgoal (or primitive action) is used to represent this causal relationship. Our Bayesian network captures knowledge about how the graphic designer's goal of conveying a message can be achieved via the viewer performing certain perceptual and cognitive tasks, as well as knowledge about how perceptual and cognitive tasks decompose into sets of simpler tasks. By *perceptual tasks*, we mean tasks that can be performed by simply viewing the graphic, such as finding the top of a bar in a bar chart; by *cognitive tasks*, we mean tasks that are done via mental computations, such as computing the difference between two numbers (Kerpedjiev and Roth 2000).

A key component of a plan inference system is the evidence that is used to guide the inference process, so that one hypothesis might eventually be preferred over another. Therefore, in extending plan inference techniques to the recognition of intentions from information graphics, we need to identify the types of evidence present in information graphics and incorporate them into our Bayesian network. Following the AutoBrief work (Kerpedjiev and Roth 2000; Green et al. 2004) on generating graphics to achieve communicative goals, we adopt the hypothesis that when constructing

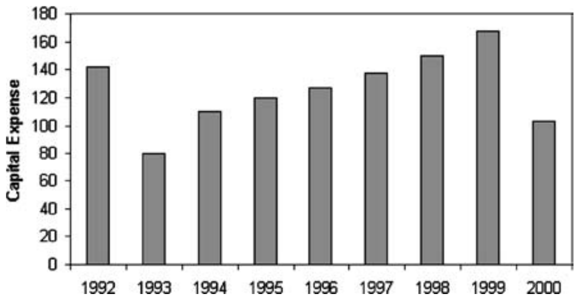


Fig. 3 Sample bar chart

the graphic, the designer made certain design decisions in order to make "important" tasks (the ones that the viewer is intended to perform in getting the graphic's message) as *salient* or as *easy* as possible. Thus tasks that are particularly salient or easy provide communicative signals as to the message that the information graphic is intended to convey.

There are a number of ways that a graphic designer can make a task salient. For example, the designer can construct a caption whose verb suggests the desired task (such as "Capital Expense Peaks in 1999" for the graphic shown in Fig. 3), where the verb *peaks* suggests that the graphic portrays the end of a trend in the data, and 1999 suggests the importance of the bar representing 1999 to the message of the graphic. The designer could also highlight specific elements in the graphic through visual cues (such as by coloring one bar in a bar chart red, while the others are black) in which case tasks involving the highlighted elements become salient.

The graphic designer can make a task *easy* for the viewer to perform by the choice of graphic type (e.g., bar chart versus line graph) (Simkin and Hastie 1987; Shah et al. 1999; Zacks and Tversky 1999; Hollands and Spence 2001) and the organization and presentation of data (Bertin 1983; Tufte 1983; Cleveland 1985; Kosslynbook 1994, among many others). For example, the task of recognizing the age group with the maximum drug spending is easier in Fig. 2b than in Fig. 2a, although both graphs depict the same data. Based on the hypothesis that the graphic designer attempts to construct a graphic that will facilitate the important tasks (making them easy for the viewer to perform), we include perceptual task effort as one of the sources of evidence in our Bayesian network for inferring the intention of an information graphic. This particular type of evidence is the focus of this paper.

The problem that we must address is how to identify the easiest perceptual tasks. Our solution is to design a model, which can rank the different perceptual tasks on a given information graphic based on the expected effort. The remainder of this paper presents our model of perceptual task effort, an eye tracking experiment that validates this model, and a brief example showing the impact of our model on our system's hypothesis about the intended message of an information graphic.

3 Developing a model of perceptual task effort

Given a set of data, the graphic designer has many alternative ways of designing a graphic. As Larkin and Simon (1987) note, information graphics that are *informationally* equivalent (all of the information in one graphic can also be inferred from the other) are not necessarily *computationally* equivalent (enabling the same inferences to be drawn quickly and easily). Peebles and Cheng (2003) further observe that even in graphics that are informationally equivalent, seemingly small changes in the design of the graphic can affect viewers' performance of graph reading tasks. Much of this can be attributed to the fact that design choices made while constructing an information graphic will facilitate some perceptual tasks more than others. Following AutoBrief, we hypothesize that the designer chooses a design that best facilitates the tasks that are most important to conveying his intended message, subject to the constraints imposed by competing tasks (Kerpedjiev and Roth 2000; Green et al. 2004).

In order to identify the perceptual tasks that the graphic designer has best enabled in the graphic, our methodology is to construct a set of rules that estimate the effort required for different perceptual tasks within a given information graphic. To develop these rules, we have applied the results of research from cognitive psychology (as described in Sect. 3.2). In doing this, we are constructing a user model representing the relative ease or difficulty, with which the viewer of a graphic could complete various perceptual tasks. The component of our system that is responsible for estimating effort is called Analysis of Perceptual Task Effort (APTE).

The goal of APTE is to determine whether a task is easy or hard to perform with respect to other perceptual tasks that could be performed on an information graphic. In order to estimate the relative effort involved in performing a task, we adopt a GOMS-like approach (Card et al. 1983) decomposing each task into a set of component tasks. Following other cognitive psychology research, we take the principal measure of the effort involved in performing a task to be the amount of time that it takes to perform the task, and our effort estimates are based on time estimates for the component tasks.[3] In this sense, our work follows that of Lohse (1993) in his UCIE system, a cognitive model of information graphic perception intended to simulate and predict human performance on graph comprehension tasks. However, we are not attempting to develop a predictive model of our own—our aim is to identify the relative difficulty of different tasks that are facilitated by the graphic designer's design choices in order to utilize that information in the plan inference process.[4] Once an initial set of rules was formulated, a preliminary eye tracking experiment was performed to verify that the cognitive principles that guided the development of the rule set were appropriately applied and to suggest modifications to individual conditions of the rules within the rule set. The resultant rules are presented in this paper, along with the methodology and results of a second eye tracking experiment used to validate them.

3.1 Structure of rules

APTE contains a set of rules that estimate how well a task is enabled in an information graphic. Each rule captures a perceptual task that can be performed on a bar chart, along with the conditions that affect the difficulty of performing that task. The conditions for the tasks are ordered so that the conditions producing the lowest estimates of effort appear first. Often several conditions within a single rule will be satisfied—this might occur, for example, in the rule shown in Fig. 4, which estimates the effort of determining the exact value represented by the top of a bar in a bar chart,

[3] The units of effort estimated by our rules roughly equate to milliseconds.

[4] Based on the goals of our model, we have adopted the more coarse-grained "Gomsian" approach rather than a very rich framework such as EPIC (EPIC 2004) or ACT-R (ACT-R 2004).

Rule-B1:Estimate effort for task
PerceiveValue(<viewer>, <g>, <att>, <e>, <v>)

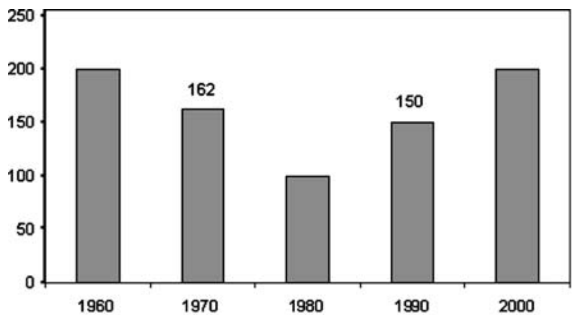
Graphic-type: bar-chart

Gloss: Compute effort for finding the exact value <v> for attribute <att> represented by top <e> of a bar <b> in graph <g>

Conditions:
B1-1: IF the top <e> of bar <b> is annotated with a value, THEN effort=150 + 300
B1-2: IF the top <e> of bar <b> aligns with a labelled tick mark on the dependent axis, THEN effort=230 + (scan + 150 + 300) x 2

Fig. 4 Rule for estimating effort for the perceptual task *Perceive Value*

Fig. 5 Information graphic illustrating conditions of rule-B1 (Fig. 4)



given that the viewer is already focused on the top of the bar. Condition–computation pair B1-1 estimates the effort involved when the bar is annotated with the value; this condition is illustrated by the second and fourth bars in Fig. 5. The second condition–computation pair, B1-2, is applicable when the top of the bar aligns with a labeled tick mark on the dependent axis; this condition is illustrated by all bars except the second bar in Fig. 5. If the top of a bar both falls on a tick mark and has its value annotated at the top of the bar (as in the fourth bar in Fig. 5), the easiest way to get the value represented by the top of the bar would be to read the annotated value, although it could also be obtained by scanning across to the tick mark. When multiple conditions are applicable, the first condition that is satisfied will be applied to calculate the effort estimate, thereby estimating the least expected effort required to perform the task.

3.2 Defining effort estimates

Researchers have examined many different perceptual tasks, although often studying individual perceptual tasks in isolation. As mentioned earlier, we have followed Lohse's approach (1993) in breaking down our tasks into component tasks. We then utilize existing time estimates (primarily those applied in Lohse's UCIE system) for the component tasks wherever possible. For some perceptual tasks, this has been a sufficient foundation for our rules. For example, we developed effort estimates for the rule shown in Fig. 6 in this manner. This rule applies a subset of the tasks given in Fig. 7 in order to estimate the effort required to read the label of a bar in a bar

Rule-B2:Estimate effort for task PerceiveLabel(<viewer>, <g>, <b>, <l>)

Graphic-type: bar-chart

Gloss: Compute effort for finding the label <l> corresponding to bar <b>
on the primary key axis of graph <g>

Conditions:
B2-1: effort=230 + 150 + 300

Fig. 6 A rule for estimating effort for the perceptual task *PerceiveLabel*

Cost	Description	Reference
1180	Interpolate on a linear scale	Boff & Lincoln (1988)
300	Recognize a six-letter word	John & Newell (1990)
230	Make a saccade	Russo (1978)
92	Make a perceptual judgment	Welford (1973)
4	Scan (per degree of visual arc)	Kosslyn (1983)
150	Discriminate an object	Based on Lohse (1993)
33	Compare two digits	Cavanaugh (1972)

Fig. 7 Cognitive engineering parameters used by UCIE

chart. In this task, the viewer is focused on the relevant bar, so the effort is estimated as 230 units for the saccade from the bar to the label (Russo 1978), 150 units for discriminating the label (based on work by Lohse 1993), and 300 units for recognizing a word (John and Newell 1990).[5] The effort estimate for this task represents one of the more straight-forward applications of the component tasks (shown in Fig. 7) in our model.

Consider, again, the rule shown in Fig. 4. The estimation of effort for the first condition, B1-1 (finding the exact value for a bar where the bar is annotated with the value), is fairly straightforward. The effort is estimated as 150 units for discriminating the label and 300 units for recognizing a 6-letter word. For the second condition, B1-2, which estimates the effort required to find the exact value represented by a bar, where the top of the bar is aligned with a tick mark on the axis, our initial effort estimate involved simply scanning over to the dependent axis (measured in terms of distance in order to estimate the degrees of visual arc scanned Kosslyn 1989) and discriminating and recognizing the label. However, our preliminary eye tracking experiments yielded some unexpected insights into the way in which participants perform this task. The patterns of eye fixations of the participants showed that when the top of the bar is aligned with a tick mark, participants frequently repeat the task, presumably to ensure accuracy, and so we modified our initial rule to include a saccade from the axis back to the top of the bar (230 units) and the repetition of the initial task (as (scan $+150+300$) $\times 2$).

Notice that *Rule-B1* (Fig. 4) does not estimate the effort required to get the value represented by the top of a bar in the case where the viewer must scan to the axis

[5] We are currently using the rough estimate of 300 units for all labels, when this value is actually the estimate of the effort required to read a 6-letter word (John and Newell 1990); this is consistent with the component tasks utilized by Lohse (1993).

Rule-B3:Estimate effort for task
PerceiveInfoToInterpolate(<viewer>, <g>, <att>, <e>, $<l_1>$, $<l_2>$, <f>)

Graphic-type: bar-chart

Gloss: Compute effort for finding the two surrounding labels $<l_1>$ and $<l_2>$ and the fraction <f> of the distance that the point representing the top <e> of a bar <b> lies between $<l_1>$ and $<l_2>$ on the axis representing attribute <att> of graph <g>, so that the value representing the top <e> of the bar can be interpolated

Conditions:
B3-1: effort=scan + 150 + (230 + 150 + 300) × 2

Fig. 8 Rule for estimating effort for the perceptual task *PerceiveInfoToInterpolate*

and interpolate an estimated value. The effort required in this case is estimated by *Rule-B3* (shown in Fig. 8). In this task, the effort estimate includes scanning over to the dependent axis, discriminating the point on the axis estimated to align with the top of the bar, along with the cost of the saccade to each surrounding label and the cost of discriminating and reading both labels. This task does not include the effort required to actually interpolate (estimate) the value represented by the point, since this is considered to be a cognitive task, and the APTE rules are intended to only represent perceptual tasks. The effort required to interpolate the value is captured in a cognitive primitive in the plan inference portion of our system.

For more complex tasks that have not been explicitly studied by cognitive psychologists, we have applied existing principles and laws in the development of our rules for estimating perceptual effort. An example of this is the class of comparison tasks (e.g., comparing the tops of two bars to determine the relative difference in value), where the proximity compatibility principle defined by Wickens and Carswell (1995) plays a major role. This principle is based on two types of proximity. *Perceptual proximity* refers to how perceptually similar two elements of a display are (in terms of spatial closeness, similar annotations, color, shape, etc.). For example, if two bars in a bar chart are colored red, while all of the other bars are colored black, the two red bars would have close perceptual proximity based on their color. *Processing proximity* refers to how closely linked the two elements are in terms of completing a particular task. If the elements must be used together (integrated) in order to complete a task, they have close processing proximity. For example, if a subset of the bars in a bar chart show a trend in the data, the bars representing the trend are closely linked in terms of processing proximity. The proximity compatibility principle states that if there is close processing proximity between two elements, then close perceptual proximity is advised. If two elements are intended to be processed independently, then distant perceptual proximity is advised. Violating the principle will increase the effort required for a viewer to process the information contained in the display.

We assume that the graphic designer attempted to follow the proximity compatibility principle in designing the information graphic so as to facilitate intended tasks and make them easier to perform than if the principle were violated. Several of our APTE rules apply the proximity compatibility principle in enumerating the conditions that affect the difficulty of performing a task—(see, e.g., the rule in Fig. 9), where the

Rule-B4: Estimate effort for task
PerceiveRelativeDifference(<viewer>, <g>, <e1>, <e2>, <b1>, <b2>, <r>)

Graphic-type: bar-chart

Gloss: Compute effort for finding the relative difference <r> in value (greater than /less than/equal to) represented by the tops <e1> and <e2> of two bars <b1> and <b2> in graph <g>

Conditions:
B4-1: IF bar <b1> and bar <b2> are adjacent and the height difference is >10% THEN effort=92 + 230 + 150
B4-2: IF bar <b1> and bar <b2> are not adjacent and the height difference is >10% THEN effort=92 + 460 + 150
B4-3: IF bar <b1> and bar <b2> have height difference >5% THEN effort=92 + 920 + 150
B4-4: ELSE effort = 92 + 1380 + 150

Fig. 9 A rule for estimating effort for perceptual task *PerceiveRelativeDifference*

effort required to perform the integrated task of determining the relative difference between two bars is different depending on the bars' spatial proximity. In particular, the effort required for adjacent bars will generally be lower than if the bars were not adjacent. We believe that this principle will also be applicable when constructing rules for estimating the effort of performing the same perceptual tasks on different types of information graphics. For example, the elements (points) in a line graph have a higher perceptual proximity than the bars in a bar chart (this example of perceptual proximity applies the Gestalt law of good continuation) (Pomerantz and Kubovy 1986). This means that it will be easier to perform integrated tasks with the points on a line in a line graph than it will be to perform the same task with the bars in a bar chart, and thus the effort estimates for these tasks on line graphs should be lower than for the same tasks on bar charts.

Many of the tasks for which we have developed effort estimates involve discriminating between two or more graphical elements; these tasks require the viewer to make comparative judgments of length, area, and angle. In order to define the conditions affecting the complexity of these judgments, we have applied Weber's Law (Cleveland 1985). One of the implications of this law is that a fixed percentage increase in line length or area is required to enable discrimination between two entities (and the probability of discrimination is affected not by the absolute difference in size, but rather by the percentage difference in size). Weber's Law has influenced the introduction of thresholds in rules such as *Rule-B4* in Fig. 9, where thresholds in the percentage difference in the height of the bars impact the effort required to perceptually discriminate the relative difference between the values represented by the bars.

In some cases, the *optimal* combination of component tasks does not take into account the escalating complexity represented by the conditions of the rule. For example, our preliminary eye tracking experiments showed that viewers performed an average of four saccades if the bars to be compared differ in height by 5–10% and an average of two saccades if the non-adjacent bars' height difference was greater than 10%. In both cases, one saccade (from the top of the lowest bar to the top of the highest bar) would be optimal in the sense of providing the necessary information. Moreover, the eye tracking experiments showed that when the height difference was small (between 5 and 10%), whether or not the bars were adjacent did not have a discernable effect on the number of required saccades. Our rules capture the expected number of saccades required by the average viewer in order to perform the necessary perceptual judgment. The effort estimates in Fig. 9 show the estimate of 92 units to perform a perceptual judgment (Welford 1973) along with a multiple of 230 units where 230 represents the estimate for a saccade (Russo 1978), and 150 units (based on Lohse 1993) in order to discriminate the top of one of the bars (usually the higher bar).

The rule for estimating the effort required to find the bar representing the maximum value is shown in Fig. 10. There is a very similar rule for finding the bar representing the minimum value in a graph (see Appendix for the complete set of rules). Conditions B5-1 and B5-2 represent a design choice that may have been made by the graph designer in order to facilitate this task. If the bars are sorted by ascending or descending order, the viewer needs only to detect this visual pattern in order to determine at which end of the graph to find the maximum value, thus eliminating the need to perform a comparison between specific bars. A graph illustrating the first condition (bars sorted by ascending height) is shown in Fig. 11. The effort estimate in this case includes the viewer scanning the graph and discriminating the top of the tallest bar.

Rule-B5: Estimate effort for task PerceiveMaximum(<viewer>, <g>, <e>, <b>)

Graphic-type: bar-chart

Gloss: Compute effort for finding the top <e> of bar <b> such that the value
represented by <e> for the attribute displayed on the dependent axis
is the maximum value in graph <g>

Conditions:
B5-1: IF the bars are sorted by ascending height THEN effort=scan + 150
B5-2: IF the bars are sorted by descending height
THEN effort=scan + 230 + 150
B5-3: IF bar <b> is more than >20% taller than next highest bar
THEN effort=scan + 460 + 150 + 92
B5-4: IF bar <b> is more than >10% taller than next highest bar
THEN effort=scan + 690 + 150 + 92
B5-5: IF bar <b> is more than >5% taller than next highest bar
THEN effort=scan + 920 + 150 + 92
B5-6: ELSE effort =scan + 1150 + 150 + 92

Fig. 10 A rule for estimating effort for perceptual task *PerceiveMaximum*

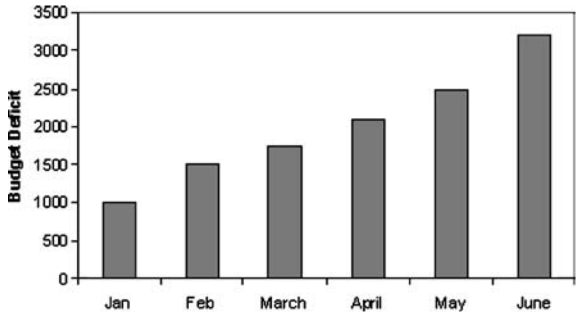


Fig. 11 Information graphic illustrating condition *B5-1* of rule-B5 (Fig. 10)

To capture the effort expended when the bars are sorted in descending height, the second condition, B5-2, includes a saccade back to the beginning of the graph in order to discriminate the tallest bar. For the remaining four conditions, Weber's Law (Cleveland 1985) has again played a critical role in the effort estimates. The effort estimates in these conditions (B5-3, B5-4, B5-5, and B5-6 in Fig. 10) differ in terms of the expected number of saccades based on various height thresholds between the maximum and the next highest bar.

In some cases, the results of our preliminary eye tracking experiment caused us to alter the component tasks applied in our effort estimates. For example, our initial APTE rules for trends represented the perception of a trend in terms of simple scans of the bar chart, while our preliminary eye fixation data showed a less smooth, slower processing of the data. By revising our APTE trend rules to represent the pairwise perceptual judgments supported by the eye fixation data (viewers tended to have half as many fixations as there were bars, with the fixation occurring between a pair of bars), we were able to more accurately assess the effort required for trend recognition. The resultant rule for calculating the effort estimates for perceiving a rising trend in a graph is shown in Fig. 12 (there are similar rules for falling and stable trends). The rule conditions vary based on the amount of variance in the trend. For example, a consistent increase in the trend is captured by the first condition. The effort required to detect a trend where the presence of one or more bars disrupt the consistency of the trend (representing a temporary drop) will be captured by the second condition. The effort estimate in the first condition, B6-1, entails the viewer making pairwise perceptual judgments (fixating on each subsequent pair of bars) across the entire graph, then making a saccade back to the beginning of the graph and making another pass of pairwise perceptual judgments over just the portion of the graph representing the trend. The scans (scan_of_graph and scan_of_trend) represent the effort required to transition between fixations. In the second condition, B6-2, the variance in the trend causes the viewer to make an additional saccade and pass over the trend, as if to verify their detection of the trend (this is supported by the preliminary eye tracking data).

Figure 13 presents our rule for estimating the effort involved in searching for a specific bar in a bar chart. The first condition is applied if the bars in the graph are displayed so that the labels are sorted in numerical or alphabetic order (as in the graph

---

Rule-B6: Estimate effort for task
PerceiveIncreasingTrend(<viewer>, <g>, <e_1>, <e_2>, <b_1>, <b_2>)

Graphic-type: bar-chart

Gloss: Compute effort for finding the tops <e_1> and <e_2> of bars <b_1> and <b_2> representing the endpoints of a rising trend in graph <g>

Conditions:
B6-1: IF there is low variance in the trend THEN
effort=ceiling($NUM_BARS_IN_CHART/2) × 92 + scan_of_graph +
230 + ceiling($NUM_BARS_IN_TREND/2) × 92 + scan_of_trend
B6-2: IF variance in the trend is less than max-allowable-variance THEN
effort=ceiling($NUM_BARS_IN_CHART/2) × 92 + scan_of_graph +
2 × (230 + ceiling($NUM_BARS_IN_TREND/2) × 92
+ scan_of_trend)

Fig. 12 A rule for estimating effort for perceptual task *PerceiveIncreasingTrend*

Rule-B7:Estimate effort for task
PerceiveBar(<viewer>, <g>, <l>, <e>, <b>)

Graphic-type: bar-chart

Gloss: Compute effort for finding the top <e> of bar <b> in graph <g>,
such that the label on the primary key axis that corresponds to <b>
is equal to <l>

Conditions:
B7-1: IF bars are sorted alphabetically or in numerical sequence
THEN effort=scan + 150 + 300 + 230
B7-2: IF bars are not sorted and bar <b> occurs before the middle bar <mid>
THEN effort=[(ceiling((mid - bar)/2) + 1) × 150 + $scan_1$ +
((ceiling(($NUM_BARS_IN_CHART - mid + 1)/2)) × 150 +
$scan_2$ + 230 +
(ceiling((bar - 1)/2) + 1) × 150 + $scan_3$)]/2 +
300 + 230
B7-3: IF bars are not sorted and bar <b> occurs at or after <mid>
THEN effort=[(ceiling((bar - mid - 1)/2) + 1) × 150 + $scan_1$ +
((ceiling(mid/2)) × 150 + $scan_2$ + 230 +
(ceiling((bar - mid - 1)/2) + 1) × 150 + $scan_3$)]/2 +
300 + 230
NOTE: <mid> is calculated as ceiling($NUM_BARS_IN_CHART/2)

Fig. 13 Rule for estimating effort for the perceptual task *PerceiveBar*

in Fig. 5 where the bars are sorted by year). The effort estimate for this condition includes the viewer scanning the bars to find the appropriate label and discriminating and reading that label before making the saccade to the top of the bar. However, our preliminary eye tracking data provided some unexpected insight into viewer behavior when labels are unsorted along the primary key axis. Our initial APTE rule for finding the top of a bar given the bar's label described the viewer as performing a left-to-right scan and sampling of the unsorted labels. In analyzing the results of our preliminary experiments, we found that viewers were far more likely to begin in the middle of the list of labels on the axis and search to the left or right, then saccade to the other half of the list of labels if they are unsuccessful in their initial search. Without being able to analyze the patterns of participants' eye fixations, we would have been unable to capture this search process and the resultant effort estimate, which is given by the formulas in conditions B7-2 and B7-3. These somewhat complicated formulas determine the average number of labels that the viewer will discriminate and reject before finding the correct bar. Because the eye tracking data indicated that viewers begin in the middle of the graph and are nearly equally as likely to go to the right as they are to go to the left in their search for the appropriate label, the effort estimates include the number of fixations in each case and take the average of the cases. The number of bars in each case is divided by two because viewers tend to fixate in between labels and generally discriminate two labels in a single fixation.[6] For example, the condition described in B7-2 is applied if the bars are unsorted and the bar being searched for falls before the midpoint of the graph. In this case, the viewer might either begin at

[6] The eye fixations included in the effort estimates of conditions B7-2 and B7-3 either fall between the labels of two bars or between the dependent axis and the label of the first bar.

the midpoint and scan left and find the bar almost immediately or the viewer might begin by scanning right (not finding the bar) and then saccade back to the beginning of the graph to scan the first half of the graph until the bar is located. Because these two scenarios appear to be nearly equally likely, the effort required for both of them is calculated, and the average is taken. The scans ($scan_1$, $scan_2$, and $scan_3$) represent the effort required to transition between fixations for each of the relevant segments of the graph. The effort estimate also includes the estimated effort required to read the appropriate label and to make the saccade to the top of the bar. B7-3 applies the same logic for the case where the bar appears either at or after the midpoint of the graph.[7]

The full set of 15 rules comprising our model of perceptual task effort for simple bar charts is shown in Appendix of this paper. This rule set encapsulates not only the cognitive principles discussed in this section but also the practical evidence gleaned from our preliminary eye tracking experiment regarding the way in which viewers go about completing the perceptual tasks being studied.

4 Validating the model

This section describes an eye tracking experiment that was conducted to evaluate the APTE rules for simple bar charts.

4.1 Method

Nineteen participants[8] were asked to perform various tasks using vertical (column) bar charts shown to them on a computer monitor while their eye fixations were recorded. Each task was completed by all of the participants. The tasks corresponded directly to the APTE rules—for example, finding the bar representing the maximum value in a bar chart (Fig. 10) and finding the exact value represented by the top of a particular bar in a bar chart (Fig. 4). For each task, participants were shown a screen with the instructions for the task displayed. The instructions for each task included some specific action to be taken by the participants to indicate the results of the task. These actions fell into two categories; in the first category, the result of the task was indicated by the participants clicking on an element of the information graphic, while results of tasks in the second category were indicated by the participants clicking on one of three buttons shown below the information graphic. For example, the task of finding the bar representing the maximum value falls into the first category, and the instructions for this task were as follows:

> In the following bar chart locate and click the top of the bar that represents the maximum value.

The task of finding the exact value represented by the top of a particular bar falls into the second category, and the instructions for this task were as follows:

> In the following bar chart there is a bar with an arrow pointing to it. Find the exact value represented by the top of that bar. Click the button that shows the correct value.

---

[7] Our preliminary eye tracking experiment showed that in this case, if viewers initially scan left and do not find the bar, they then tend to saccade back to the middle to continue the search.

[8] Three additional participants could not be calibrated on the eye tracker. Also, note that the participants involved in this experiment were different from the subjects used in our preliminary eye tracking experiment.

Both categories of tasks included a mouse movement and a mouse click, and the time of the start of the mouse movement and the time and location of the mouse click were recorded as part of the data collected during the experiment. When the participants had read and felt that they understood the instructions, they clicked the mouse. The next screen that the participants were shown contained only a fixation point. After clicking on the fixation point, the participants were shown the bar chart on which they were to perform the prescribed task. The participants moved to the next task by clicking on a "Done" button shown at the bottom right corner of the screen.

4.2 Design

The experiment was designed to obtain the average time required to complete a given task across participants. Six base bar charts were constructed that displayed a variety of different characteristics (increasing versus decreasing trends, varying numbers of bars, sorted versus unsorted labels, bars sorted by height or unsorted, etc.). The six base bar charts[9] used in this experiment are shown in Fig. 14. Three of the bar charts (shown in Fig. 14 a–c) were used in the preliminary eye tracking experiment. The bar charts shown in Fig. 14d–f were designed for this experiment and had not been used previously. Note that the bar charts are fairly uniform in design—we did not test the impact of color, patterns, or backgrounds on user performance, although this would be an interesting area for future work.

The APTE rule set for bar charts currently contains 15 rules describing various perceptual tasks that can be performed. For a given bar chart, a set of between 11 and 14 tasks was tested. Not all rules can be successfully applied to all graphics. For example, it is not possible to detect a rising trend in Fig. 14a. However, other rules might have multiple conditions that could all apply to a given base bar chart [e.g., finding the relative difference (Fig. 9) is possible with many different pairs of bars]. The task sets were designed to test a wide variety of rules and conditions. For example, the tasks examined for one of the bar charts (Fig. 14a) include:

1. PerceiveValue (condition 1),
2. PerceiveValue (condition 2),
3. PerceiveIsMinimum (condition 2),
4. PerceiveLabel,
5. PerceiveRelativeDifference (condition 2),
6. PerceiveRelativeDifference (condition 4),
7. PerceiveBar (condition 1) (get bar labeled 1998),
8. PerceiveBar (condition 1) (get bar labeled 2002),
9. PerceiveIsMaximum (condition 6),
10. PerceiveMaximum (condition 5),
11. PerceiveMinimum (condition 3),
12. PerceiveInfoToInterpolate,
13. PerceiveDecreasingTrend (condition 2).

In order to prevent participants from becoming familiar with the bar charts being analyzed, the actual test graphics were variants of these "base" bar charts. In designing the test graphics, characteristics of the base bar chart that were extraneous to the

[9] We call these the "base" bar charts, since the actual bar charts used in the experiment were variants of these.

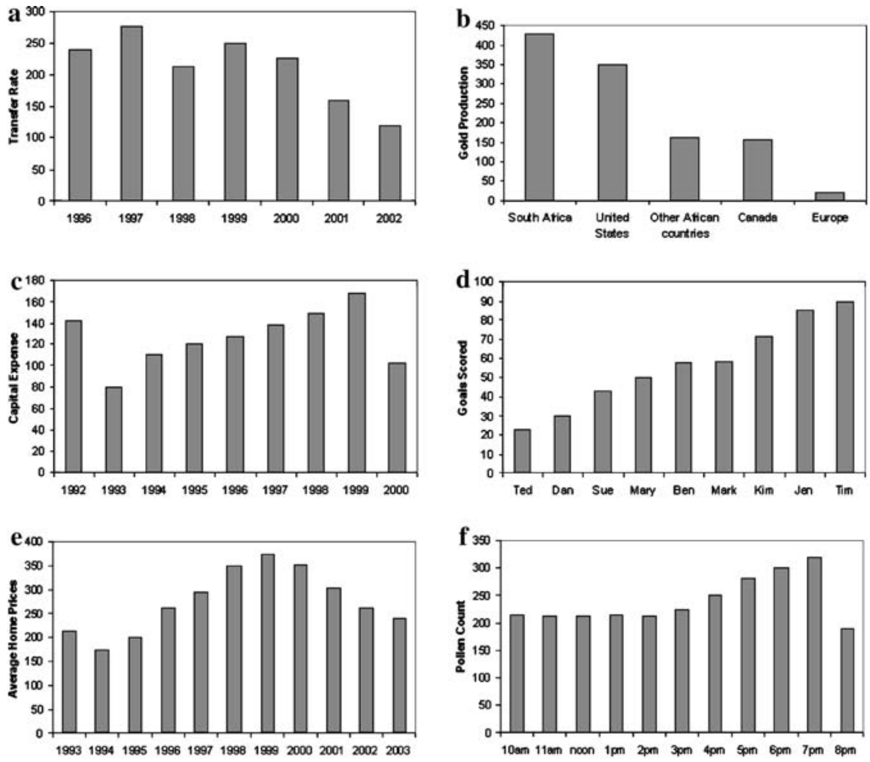


Fig. 14 Base bar charts used in eye tracking experiment

task being evaluated were altered. For example, if the task was to locate and read the label of a given bar in the base bar chart, the attribute name displayed on the *y*-axis and the heights of the bars not involved in the task would be altered in the test graphic (see Fig. 15b). If the task was to locate the top of a bar given the label of that bar, then the heights of the bars might be varied but the labels would all be consistent with the base chart (see Fig. 15c). For tasks such as trends or perceiving the minimum or maximum in a bar chart, all of the bar heights would remain the same as in the base chart, although the labels of the bars and of the dependent axis would change (see Fig. 15d).[10] The order in which the participants completed the tasks was also varied so as to further mitigate the effects of familiarity with the content of the specific information graphics and the impact of expertise obtained through practice in performing the requested tasks.

4.3 Procedure

Each trial began with a series of five practice tasks. Participants were informed that the first five tasks were for practice, and were allowed to ask any questions they had about the format of the experiment during the warmup period. Participants were given several points of instruction. For tasks that required participants to choose an answer

[10] In some cases where the values displayed along the dependent axis were extraneous to the task, the values were not displayed as in Fig. (15c, d).

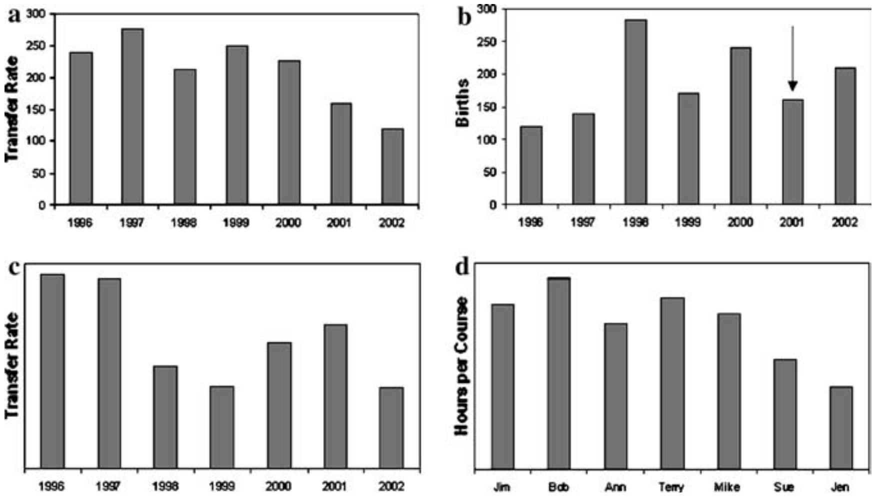


Fig. 15 A base bar chart and test variants **a** Base bar chart. **b** Chart used for perceive-label. **c** Chart used for perceive-bar (1998). **d** Chart used for perceive-maximum

shown on one of three response buttons, participants were instructed to look only at the graphic in completing the task, and to determine the answer to the task before reading the labels on the buttons. For all tasks, participants were asked to only move the mouse when they were ready to make a response. After the fifth practice graphic, the participants were presented with the series of tasks comprising the experiment. The participants' eye fixations were measured using a Tobii Model ET-17 eye tracking processor, which samples the eye every 20 ms. Individual samples were grouped into fixations when three or more consecutive samples occurred within 1° of visual angle.

4.4 Data analysis

The aim of this experiment was to obtain the average completion time of all participants for a given task, and to compare the rank order of those average completion times to the rank order of estimates produced by the APTE rules for those same tasks. The completion time for each set of data was determined based on a combination of the time of the initial mouse movement and the pattern of the participant's eye fixations. For tasks that required the user to click on a button, task completion time was recorded as the beginning of the mouse movement if that movement was just prior to the participant moving his gaze to the region of the screen where the buttons were located. If the mouse movement did not coincide with the shift in gaze away from the graphic and toward the buttons, the task completion time was recorded as the end time of the final fixation within the information graphic. For tasks that required the participant to click on an element of the information graphic, task completion time was recorded as the beginning of the mouse movement if that movement took place during the same fixation in which the participant "clicked" on the appropriate element. If the mouse movement did not coincide with this eye fixation, the completion time of the task was recorded as the beginning of the fixation during which the participant selected the appropriate graphical element. Data was excluded from the results if the participant's gaze left the information graphic to view the labels on

Task	Rule Condition	Rank from APTE Rules	Actual Rank
PerceiveIsMinimum	2	1	2
PerceiveValue	1	2	1
PerceiveLabel	1	3	3
PerceiveBar	1	4	5
PerceiveRelativeDifference	2	5	4
PerceiveBar	1	6	7
PerceiveMinimum	3	7	9
PerceiveIsMaximum	6	8	6
PerceiveValue	2	9	11
PerceiveMaximum	5	10	8
PerceiveDecreasingTrend	2	11	12
PerceiveInfoToInterpolate	1	12	10
PerceiveRelativeDifference	4	13	13

Fig. 16 Rankings for second base bar chart task set

the buttons before completing the task, if the participant responded incorrectly to the task, or if it was clear that the participant was performing processing not required by the task.

Each participant completed a total of 81 tasks, including the five "practice" tasks and the 11–14 tasks for each of the six base bar charts being examined. The set of tasks for each base bar chart was evaluated separately, since our goal is to predict the relative ease or difficulty of completing a variety of tasks on a given graphic. For each individual task, the APTE effort estimate was calculated, as was the average task completion time (over all participants) for that task.[11] The set of tasks for each base bar chart was sorted by the average task completion time for each task. These sorted task lists were compared to the sorted lists of the effort estimates produced by the APTE rules. For example, Fig. 16 shows the tasks evaluated for the base bar chart shown in Fig. 14a along with the ranking of the effort estimates generated by the APTE rules, and the ranking of the tasks in terms of the actual effort required.

4.5 Results

Given the actual and estimated rankings for the tasks for each base bar chart, we performed a statistical analysis to test the correlation between the average completion times and the effort estimates produced by our rules. We performed a Spearman Rank-Order Correlation (rho), used to determine whether two sets of rank-ordered data are related, since we are primarily interested in ranking the effort estimates generated by APTE. In order to obtain a measure of the effectiveness of our rules, we also compared our results against a baseline condition. This baseline was calculated by estimating the effort required to complete each task using only the "easiest" (or first) conditions of the rules. This gives us an idea of what the correlation between the actual rankings and the estimated rankings would be if our rules did not take into account

[11] The actual task completion times for the various participants showed a fairly wide range of values, with nearly every task containing one or two outlying values. Thus it is not surprising that the average standard deviation per task was 679 ms. However, what we are really interested in is the relative difficulty of the different tasks.

Graph	Tasks	APTE Correlation	p-value	Baseline Correlation	p-value
—–	—–	—–-	—–-	—––	—–-
a	13	0.93*	0.0001	0.21	0.4912
b	13	0.74*	0.0041	0.33	0.2692
c	11	0.70*	0.0165	0.22	0.5193
d	12	0.83*	0.0008	0.58*	0.0479
e	11	0.64*	0.0353	0.44	0.1797
f	13	0.74*	0.0098	0.10	0.7699

∗ These correlations are statistically significant at the .05 level
(using a 2-sided T-test)

Fig. 17 Correlations between average completion time and APTE effort estimates

the design choices that can make a task more difficult to perform. Figure 17 shows the results of the correlations between the average completion times for our human subjects and our APTE effort estimates for each of the six base bar charts (values approaching 1 show a strong correlation), as well as the correlation between the average completion times and the baseline condition. The *p*-values show the probability of such a correlation occurring by chance.

The results of the Spearman Rank-Order Correlation show a significant correlation for the APTE rules on each of the six graphs. Each correlation between the average task completion times and the APTE estimates strongly outperforms the correlation between the average task completion times and the baseline represented by the easiest task conditions. For the baseline, only the correlation for one base bar chart (shown in Fig. 14d) is statistically significant. The fact that this baseline correlation is statistically significant is not surprising, due to the fact that a large number of tasks in this task set involved applying the first condition of the relevant rules.

5 Discussion and future work

The results from the eye tracking experiment described in the previous section validated the rankings of perceptual task effort for simple bar charts produced by our set of APTE rules. This section discusses some of the implications of the assumptions underlying our model, as well as detailing some areas for potential future work.

5.1 Implications of our model of perceptual task effort

As described previously, our use of the APTE rules reflects our hypothesis that since the graphic designer has many alternative ways of designing a graphic, the designer chooses a design that best facilitates the tasks that are most important to conveying his intended message, subject to the constraints imposed by competing tasks (Kerpedjiev and Roth 2000). Underlying this hypothesis is the assumption that the graphic designer is competent, and that a competent graphic designer has a fairly accurate model of what is required to perceptually facilitate the tasks. We have based this assumption on the wealth of resources describing ways in which graphic designers can and should facilitate tasks for their viewers (e.g., Tufte 1983; Cleveland 1985; Kosslyn 1994), and the observation that many of the techniques described in these resources correspond to the cognitive principles upon which we based our APTE rules. Two issues arise from our use of this cognitive model of perceptual task effort in recognizing the intention underlying an information graphic.

The first issue is that it is unclear that graphic designers would have an accurate model of some of the less expected results, such as the similarity in effort between reading a value from a tick mark and gathering the information necessary to interpolate the value. It seems reasonable that a graphic designer wishing to facilitate the task of determining the exact value would align the top of the bar with a tick mark rather than forcing the viewer to interpolate the value. This problem of what to represent in the APTE rules accentuates the distinction between the actual effort, which viewers expend in performing particular tasks versus the difficulty the viewer might reasonably be expected to have in performing the task. Viewers tend to repeat the process of locating, discriminating, and reading the value on a tick mark, but the expected difficulty of that task would not include this repetition. Another example in which the data does not conform to expectations is the task of finding a particular label amongst the unsorted labels along the primary key axis of the bar chart. A graphic designer might reasonably place a bar first on the axis, expecting that this will reduce the difficulty of locating the bar (expecting the viewer to use a left-to-right search). In practice, it seems that viewers begin in the middle of the axis, so the task might actually be better facilitated by placing the bar in the middle of the bar chart.

In order to base the APTE rules on the expected difficulty of the tasks rather than the actual effort required to complete them, we would need to have evidence of what is contained in the stereotypical graphic designer's model of expected difficulty. Up to this point, we have been using the resources and guidelines published for graphic designers along with the assumption of competence of the graphic designer in order to draw reasonable conclusions about the knowledge of the stereotypical graphic designer. Unfortunately, the resources that we have examined do not provide guidelines for the tasks in question. Lacking evidence to support this intuitive distinction between expected difficulty and the actual effort expended in performing tasks, we have based our APTE rules on the solid evidence regarding the effort expended by participants performing these tasks that we collected during our preliminary eye tracking experiments. Therefore, what is represented by our APTE rules is a model of the effort required by a stereotypical viewer to complete the various perceptual tasks on a given information graphic.

The second issue relates to poorly designed graphics. Our goal is to infer the intended message of an information graphic. The underlying hypothesis of our work is that the graphic designer will construct a graphic that facilitates the important perceptual tasks and thus that our estimates of perceptual effort can help hypothesize those tasks and consequently contribute to recognizing the designer's intended message. But what if the graphic designer is incompetent and has produced a poorly designed graphic? In that case, the model will still provide evidence about which perceptual tasks are enabled (whether or not those are the tasks that the designer intended to enable) and this will provide evidence that will lead to what most viewers of the graphic would infer as the intended message of the graphic designer.

These issues link directly to two interesting directions for future research. The first would be to investigate the difference between expected difficulty and actual effort and its influence on the choices made by graphic designers. In order to do this, we would involve graphic designers in the research so that we could directly examine the intended message of the graphic and the design choices that are made. The other direction of research would involve improving the design choices that are made by graphic designers. Suboptimal graphic design is a common and persistent problem (see, e.g., Tufte 1983; Cleveland 1985, among many others). By using our model to provide direct feedback to the graphic designer about the tasks that are perceptually enabled in a graphic, the designer could verify that the tasks that the viewer is intended to perform are the ones being enabled.

5.2 Potential rule modifications

The results from the eye tracking experiment described in this paper validated that our model of perceptual task effort does enable us to successfully rank the relative effort required to perform various perceptual tasks on a given information graphic. However, we noticed several possible modifications that might further improve the model. We believe that these modifications would have less impact than the modifications made after our preliminary eye tracking experiment, and that some measure of iterative improvement in the model will always be possible.

The most important of the modifications that could be made would be to consider the effect of disruptions in the viewer's sightline. In bar charts, for example, the height of the intervening bars between the elements involved in a task seems to have an impact on the effort required to perform the task. For example, consider the task of perceiving the relative difference between two bars. Our rule (Fig. 9) takes into account whether or not the two bars are consecutively located, and the threshold height differences between the two bars, but does not consider the height of the bars in between the two bars being compared. The eye tracking data that we have gathered shows that the height of these intervening bars can impact the effort required to perform the task. The bar chart shown in Fig. 18 was used to test condition B4-2 since the bars are not consecutive, but do have a height difference greater than 10%. In our analysis, this task was more difficult for viewers to perform than our APTE rules estimated that it would be. This same concept also arises in several of our other rules. For example, when finding the bar representing the maximum (or minimum) of a bar chart, if the viewer is comparing two bars in order to determine which is the maximum, and the sightline between the two bars is disrupted, the task is likely to be more difficult than if the sightline was unobstructed. Intervening bars can also impact the task of getting the exact value represented by the top of a bar when the top of the bar aligns with a tick mark on the dependent axis (rule in Fig. 4). For this task, we

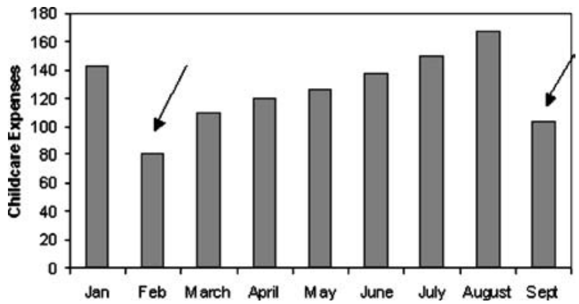


Fig. 18 Information graphic illustrating condition B4-2 of rule-B4 (Fig. 9)

have observed that the viewer is more likely to double-check the value if there are intervening bars obstructing the view of the axis.

Similarly, rules could be modified to take into account the impact of gridlines in the bar chart. Some bar charts contain horizontal gridlines that demarcate particular units along the dependent axis. By providing a visual reference point, these gridlines may facilitate certain tasks, such as finding the exact value represented by the top of a bar or finding the information needed to interpolate the value. Gridlines might also work to mitigate the effect of intervening bars on the effort required to perform some of the tasks described earlier.

There are also a few modifications that could be made to specific rules, which we would expect to improve our results. One of these involves a modification to the first two conditions of PerceiveMaximum (Fig. 10) and PerceiveMinimum. These conditions apply to finding the top of the bar representing the maximum (or minimum) value in the graph when the bars are displayed in order of either ascending or descending height, as in Fig. 14b, d. While the scan path is as described by the effort estimates (a single scan of the graph in the case of finding the maximum when the bars are displayed by ascending height, or a scan with a saccade when finding the maximum if the bars are displayed by descending height), the viewer seems to be performing some additional processing at the end of the task. There is often a notable delay between when the viewer fixates on the maximum and when they actually click the mouse to respond to the task. It is unclear at this point what this processing entails, or whether it should be included in the effort estimate for the task, but it is an interesting issue for future research.

Our current model captures the relative effort that will be required for a *stereotypical* viewer to perform a variety of perceptual tasks. An interesting issue for future work is investigating ways of tailoring our model to the expected audience for an information graphic. For example, Shah has shown that inferences that a viewer draws from a graph can be impacted by their graph reading skill (Shah 2002). Assuming that the graphic designer takes the anticipated audience and their abilities into account in designing the graphic, this means that the expected effort estimates for tasks in our model should perhaps be different based on the intended audience for the graph. So a specific task could have one effort estimate if the graphic appears in a graduate-level textbook, and a different (harder) estimate if that graphic appears in a middle-school textbook.

Although we have listed a number of possible future modifications to the APTE rules, many of which have been suggested by an examination of the eye tracking data we have collected, it is important to note that the current model does, in fact, enable us to rank the expected effort required for different perceptual tasks. This has been demonstrated in the significant correlations that we have obtained between the rank order of average completion times of viewers and the rank order of our estimated effort for those same tasks. The next section gives a brief overview of how our model of perceptual effort is used in hypothesizing the message conveyed by a simple bar chart.

6 Exploiting perceptual task effort in our bayesian network: a brief overview

The APTE rules play two roles in our graph understanding system. If all possible perceptual tasks were used to build the Bayesian network, the resultant network would be huge and would exceed memory limits. Thus we first identify the easiest perceptual tasks for the particular graphic, with at most one instantiation of each generic task. For example, we would not want to solely use instantiations of the PerceiveLabel task. These easiest perceptual tasks, along with other tasks that are suggested by other features of the graphic such as salience, are used as the starting point for building the Bayesian network. The second way in which the APTE rules are used is to provide evidence about which tasks the viewer was intended to perform, under the assumption that the graphic designer tried to facilitate those tasks that were important in gleaning the graphic's intended message. The conditional probability tables in the Bayesian network capture the relationship between the relative effort of a perceptual task and the perceptual task being part of the plan for understanding the graphic's message.[12]

To see the impact of relative effort on hypothesizing a graphic's intended message, consider the graphic that appears in Fig. 19. Our system hypothesizes that the graphic is intended to convey that the US ranks third among the countries shown in the bar chart with respect to GDP per capita and assigns this intention a likelihood of more than 95%. Note that the task of perceiving the rank of the US is relatively easy due to the bars appearing in order of height. Now consider a significant variation of the graphic design. Suppose that the bars were sorted by the alphabetical order of their labels, rather than by descending order of their height. This variation is shown in Fig. 20. The perceptual task of determining the rank of the US is now estimated as

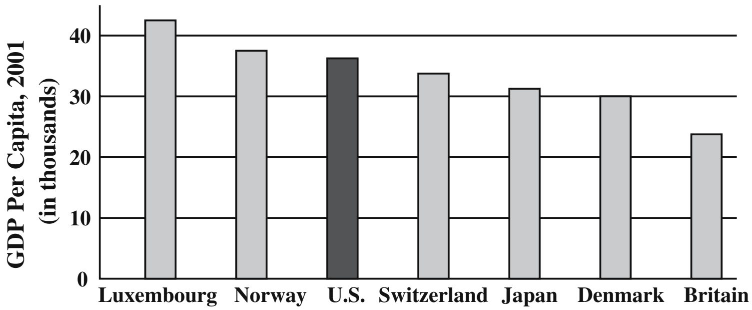


Fig. 19 Information graphic example with bars sorted by height

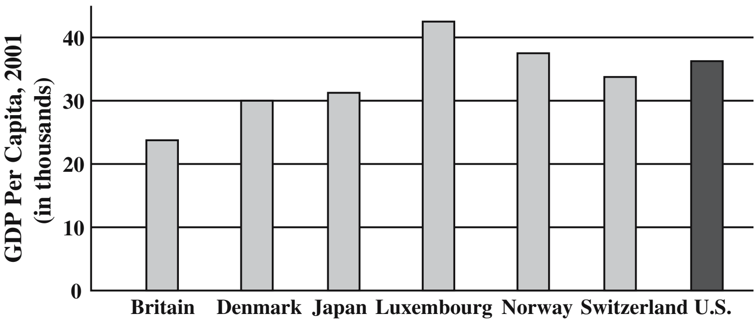


Fig. 20 Information graphic example with bars sorted by label

[12] The interested reader should see Elzer et al. 2005 for more details on our Bayesian network.

being difficult. This new evidence results in the system assigning a probability of less than 10% to the GetRank message.

7 Conclusion

This paper has presented a model of perceptual task effort for simple bar charts. We have presented the theoretical underpinnings of the model as well as refinements suggested by actual observation of viewers during preliminary eye tracking experiments. The results of a subsequent eye tracking experiment demonstrate the ability of our model to successfully rank the difficulty of perceptual tasks on an information graphic according to the amount of effort required to complete those tasks. In addition, we have outlined the role that this model of the user plays in recognizing the graphic designer's intended message for an information graphic. The ability to infer the intended message of a graphic will play a vital role in the summarization of information graphics in order to (1) provide access to publications in digital libraries via the content of information graphics, (2) provide alternative access to information graphics for visually impaired viewers, and (3) provide alternative access to information graphics in environments with limited bandwidth or miniature viewing facilities.

Acknowledgements This work is being partially supported by the National Science Foundation under Grant No. IIS-0534948. The work of the second author was partially supported by the National Science Foundation under CAREER Award No. 0132821.

Appendix

APTE rules

Rule-B1 Estimate effort for task PerceiveValue:
Given the top of bar <b>, compute effort for finding the exact value <v> for attribute <att> represented by top <e> of a bar <b> in graph <g>
B1-1: IF <b> is annotated with value, THEN effort = 150 + 300
B1-2: IF <b> occurs on a tick mark annotated with value,
THEN effort = 230 + (scan + 150 + 300) × 2

Rule-B2 Estimate effort for task PerceiveLabel:
Given the top of bar <b>, compute effort for finding the label <l> corresponding to bar <b> on the primary key axis of graph <g>
B2-1: effort = 230 + 150 + 300

Rule-B3 Estimate effort for task PerceiveInfoToInterpolate:
Given the top of bar <b>, compute effort for finding the two surrounding labels, <l_1> and <l_2>, and the fraction <f> of the distance that the point representing the top <e> of a bar <b> lies between <l_1> and <l_2> on the axis representing attribute <att> of graph <g>, so that the value representing the top <e> of the bar can be interpolated
B3-1: IF the axis is labelled with values,
THEN effort = scan + 150 + (230 + 150 + 300) × 2

Rule-B4 Estimate effort for task PerceiveRelativeDifference:
Given the tops of bars <b1> and <b2>, compute effort for finding the relative difference <r> in value (greater than/less than/equal to) represented by the tops <e1> and <e2> of two bars <b1> and <b2> in graph <g>

B4-1: IF <b1> and <b2> are consecutively located and difference >10%,
THEN effort = 92 + 230 + 150
B4-2: IF <b1> and <b2> are not consecutive and difference >10%,
THEN effort = 92 + 460 + 150
B4-3: IF <b1> and <b2> have difference >5%, THEN effort = 92 + 920 + 150
B4-4: ELSE effort = 92 + 1380 + 150

Rule-B5 Estimate effort for task PerceiveMaximum:
Compute the effort for finding the top <e> of bar <b> such that the value represented by <e> for the attribute displayed on the dependent axis is the maximum value in graph <g>

B5-1: IF the bars are sorted by ascending height, THEN effort = scan + 150
B5-2: IF the bars are sorted by descending height, THEN effort = scan + 150 + 230
B5-3: IF <b> is more than 20% taller than next highest bar,
THEN effort = scan + 460 + 150 + 92
B5-4: IF <b> is more than 10% taller than next highest bar,
THEN effort = scan + 690 + 150 + 92
B5-5: IF <b> is more than 5% taller than next highest bar,
THEN effort = scan + 920 + 150 + 92
B5-6: ELSE effort = scan + 1150 + 150 + 92

Rule-B6 Estimate effort for task PerceiveIncreasingTrend:
Compute effort for finding the tops <e_1> and <e_2> of bars <b_1> and <b_2> representing the endpoints of a rising trend in graph <g>

B6-1: IF low variance, THEN
effort = (ceiling(bars/2)) × 92 + 230 + (ceiling(bars-in-seg/2)) × 92
B6-2: IF higher but still acceptable variance, THEN
effort = (ceiling(bars/2)) × 92 + 2 × (230 + ceiling(bars-in-seg/2)) × 92

Rule-B7 Estimate effort for task PerceiveBar:
Given a label <l>, compute effort for finding the top <e> of bar <b> in graph <g>, such that the label on the primary key axis that corresponds to <b> is equal to <l>

B7-1: IF bars are sorted alphabetically or in numerical sequence, THEN
effort = scan + 150 + 300 + 230
B7-2: IF bars are not sorted and bar B < mid (mid = ceiling(totbars/2)), THEN
effort = (((ceiling((mid − bar)/2) + 1) × 150) +
((ceiling((totbars − mid + 1)/2)) × 150 + 230 +
(ceiling((bar − 1)/2) + 1) × 150))/2 + 300 + 230
B7-3: IF bars are not sorted and bar B >= mid, THEN
effort=(((ceiling((bar − mid − 1)/2)+ 1) × 150)+(ceiling(mid/2)×150 +
230 + ((ceiling((bar − mid − 1)/2) + 1) × 150)))/2 + 300 + 230

Rule-B8 Estimate effort for task PerceiveMinimum:
Compute the effort for finding the top <e> of bar <b> such that the value represented by <e> for the attribute displayed on the dependent axis is the minimum value in graph <g>

B8-1: IF the bars are sorted by descending height, THEN effort = scan + 150
B8-2: IF the bars are sorted by ascending height, THEN effort = scan + 230 + 150
B8-3: IF <b> is more than 20% shorter than next smallest value, THEN effort = scan + 460 + 150 + 92
B8-4: IF <b> is more than 10% shorter than next smallest value, THEN effort = scan + 690 + 150 + 92
B8-5: IF <b> is more than 5% shorter than next smallest value, THEN effort = scan + 920 + 150 + 92
B8-6: ELSE effort = scan + 1150 + 150 + 92

Rule-B9 Estimate effort for task PerceiveDecreasingTrend:
Compute effort for finding the tops $<e_1>$ and $<e_2>$ of bars $<b_1>$ and $<b_2>$ representing the endpoints of a falling trend in graph <g>

B9-1: IF low variance, THEN effort = (ceiling(bars/2)) × 92 + 230 + (ceiling(bars-in-seg/2)) × 92
B9-2: IF higher but still acceptable variance, THEN effort = (ceiling(bars/2)) × 92 + 2 × (230 + ceiling(bars-in-seg/2)) × 92

Rule-B10 Estimate effort for task PerceiveStableTrend:
Compute effort for finding the tops $<e_1>$ and $<e_2>$ of bars $<b_1>$ and $<b_2>$ representing the endpoints of a stable trend in graph <g>

B10-1: IF low variance, THEN effort = (ceiling(bars/2)) × 92 + 230 + (ceiling(bars-in-seg/2)) × 92
B10-2: IF higher but still acceptable variance, THEN effort = (ceiling(bars/2)) × 92 + 2 × (230 + ceiling(bars-in-seg/2)) × 92

Rule-B11 Estimate effort for task PerceiveIfSorted:
Compute effort for determining whether the bars are displayed in sorted (ascending or descending) order of height along the primary key axis

B11-1: effort = scan of graph

Rule-B12 Estimate effort for task PerceiveRank:
Given the top of bar <b>, compute effort for finding the rank of the value represented by the top <e> of <b> on the dependent axis relative to the values represented by the tops of all of the bars displayed in the graph <g>

B12-1: IF the bars are sorted by ascending or descending height, THEN effort = 230 + scan + 150 × rank
B12-2: IF bars are not sorted, THEN effort = (combinations(bars,2) × (92 + 150 + 230)) + ((bars − 1) × 230)

Rule-B13 Estimate effort for task PerceiveChangeTrend:
Given the top <e> of bar <b>, compute effort for recognizing that the trend in the attribute displayed on the dependent axis changes at <b>

B13-1: IF there is a large change in slope, THEN effort = 230 + scan trends + 230 + 92

B13-2: IF there is a small change in slope,
THEN effort = 2 × (230 + scan trends + 230) + 92

Rule-B14 Estimate effort for task PerceiveIsMaximum:
Given the top <e> of bar <b>, compute effort for determining that <b> represents the maximum value in graph <g> for the attribute <att> displayed on the dependent axis

B14-1: IF the bars are sorted by ascending or descending height and <b> is an endpoint, THEN effort = scan of graph + 230
B14-2: IF bars are not sorted, <b> is an endpoint and <b> is more than 10% taller than next highest bar, THEN effort = scan of graph + 230 + 92
B14-3: IF bars are not sorted, <b> is an endpoint and <b> is more than 5% taller than next highest bar, THEN effort = scan of graph + 460 + 92
B14-4: IF bars are not sorted, <b> is not an endpoint and <b> is more than 10% taller than next highest bar, THEN effort = scan of graph + 460 + 92
B14-5: IF bars are not sorted, <b> is an endpoint and <b> is 5% or less taller than next highest bar, THEN effort = scan of graph + 690 + 92
B14-6: IF bars are not sorted, <b> is not an endpoint and <b> is more than 5% taller than next highest bar, THEN effort = scan of graph + 920 + 92
B14-7: IF bars are not sorted, <b> is not an endpoint and <b> is 5% or less taller than next highest bar, THEN effort = scan of graph + 1150 + 92

Rule-B15 Estimate effort for task PerceiveIsMinimum:
Given the top <e> of bar <b>, compute effort for determining that <b> represents the minimum value in graph <g> for the attribute <att> displayed on the dependent axis

B15-1: IF the bars are sorted by ascending or descending height and <b> is an endpoint, THEN effort = scan of graph + 230
B15-2: IF bars are not sorted, <b> is an endpoint and <b> is more than 10% shorter than next highest bar, THEN effort = scan of graph + 230 + 92
B15-3: IF bars are not sorted, <b> is an endpoint and <b> is more than 5% shorter than next highest bar, THEN effort = scan of graph + 460 + 92
B15-4: IF bars are not sorted, <b> is not an endpoint and <b> is more than 10% shorter than next highest bar, THEN effort = scan of graph + 460 + 92
B15-5: IF bars are not sorted, <b> is an endpoint and <b> is 5% or less shorter than next highest bar, THEN effort = scan of graph + 690 + 92
B15-6: IF bars are not sorted, <b> is not an endpoint and <b> is more than 5% shorter than next highest bar, THEN effort = scan of graph + 920 + 92
B15-7: IF bars are not sorted, <b> is not an endpoint and <b> is 5% or less shorter than next highest bar, THEN effort = scan of graph + 1150 + 92

References

ACT-R: 'The ACT-R Home Page'. http://act-r.psy.cmu.edu/ (retrieved on May 18th, 2004)
Bertin, J.: Semiology of Graphics. The University of Wisconsin Press, Madison, WI (1983)
Boff, K.R., Lincoln, J.E.: Engineering Data Compendium: Human Perception and Performance. AAMRL, Wright-Patterson AFB, OH (1988)
Carberry, S.: Techniques for plan recognition. User Model. User-Adap. Interact. **11**(1–2), 31–48 (2001)

Card, S.K., Moran, T.P., Newell, A.: The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates Inc., Hillsdale, NJ (1983)
Cavanaugh, J.P.: Relation between the intermediate memory span and the memory search rate. Psychol. Rev. **79**, 525–530 (1972)
Charniak, E., Goldman, R.P.: A Bayesian model of plan recognition. Artif. Intell. **64**(1), 53–79 (1993)
Clark, H.: Using Language. Cambridge University Press, Cambridge (1996)
Cleveland, W.S.: The Elements of Graphing Data. Chapman and Hall, New York (1985)
Corio, M., Lapalme, G.: Generation of texts for information graphics. In: Proceedings of the 7th European Workshop on Natural Language Generation EWNLG'99. pp. 49–58 (1999)
Druzdzel, M.J., van der Gaag, L.C.: Building probabilistic networks: Where do the numbers come from? IEEE Trans. Knowl. Data. Eng. **12**, 481–486 (2000)
Elzer, S., Carberry, S., Zukerman, I., Chester, D., Green, N., Demir, S.: A probabilistic framework for recognizing intention in information graphics. In: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI). pp. 1042–1047 (2005)
EPIC: The brain, cognition, and action laboratory: EPIC. http://www.umich.edu/ bcalab/epic.html (retrieved on May 18th, 2004)
Futrelle, R.: Summarization of diagrams in documents. In: Mani, I., Maybury, M. (eds.) Advances in Automated Text Summarization, pp. 403–421 MIT Press, Cambridge (1999)
Futrelle, R., Nikolakis, N.: Efficient analysis of complex diagrams using constraint-based parsing. In: Proceedings of the Third International Conference on Document Analysis and Recognition. pp. 782–790 (1995)
Green, N., Carenini, G., Kerpedjiev, S., Mattis, J., Moore, J., Roth, S.: Autobrief: an experimental system for the automatic generation of briefings in integrated text and information graphics. Int. J. Hum-Comp. Stud. **61**(1), 32–70 (2004)
Grice, H.P.: Utterer's meaning and intentions. Philos. Rev. **78**, 147–177 (1969)
Hollands, J.G., Spence, I.: The discrimination of graphical elements. Appl. Cogn. Psychol. **15**, 413–431 (2001)
Iverson, G., Gergen, M.: Statistics: The Conceptual Approach. Springer-Verlag, New York (1997)
John, B.E., Newell, A.: Toward an engineering model of stimulus response compatibility. In: Gilmore, R.W., Reeve, T.G. (eds.) Stimulus-Response Compatibility: An Integrated Approach. pp. 107–115. North-Holland, New York (1990)
Kerpedjiev, S., Roth, S.F.: Mapping communicative goals into conceptual tasks to generate graphics in discourse. In: Proceedings of Intelligent User Interfaces. pp. 157–164 (2000)
Kosslyn, S.: Elements of Graph Design. W. H. Freeman and Company, NY (1994)
Kosslyn, S.M.: Ghosts in the Mind's Machine. Norton, New York (1983)
Kosslyn, S.M.: Understanding charts and graphs. Appl. Cogn. Psychol. **3**, 185–226 (1989)
Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth a thousand words. Cogn. Sci. **11**, 65–99 (1987)
Lohse, G.L.: A cognitive model for understanding graphical perception. Hum-Comp. Interact. **8**, 353–388 (1993)
Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, CA (1988)
Peebles, D., Cheng, P.C.-H.: Modeling the effect of task and graphical representation on response latency in a graph reading task. Hum. Factors. **45**, 28–46 (2003)
Perrault, C.R., Allen, J.F.: A plan-based analysis of indirect speech acts. Am. J. Comput. Ling. **6**(3–4), 167–182 (1980)
Pomerantz, J.R., Kubovy, M.: Theoretical approaches to perceptual organization. In: Boff, K.R., Kaufman, L., Thomas, J.P. (eds.) Handbook of Perception and Human Performance, pp. 36.1–36.46. Wiley, New York (1986)
Russo, J.E.: Adaptation of cognitive processes to eye movement systems. In: Senders, J.W., Fisher, D.F., Monty, R.A. (eds.) Eye Movements and Higher Psychological Functions. Lawrence Erlbaum Associates Inc., Hillsdale, NJ (1978)
Shah, P.: Graph comprehension: the role of format, content, and individual differences. In: Anderson, M., Meyer, M.B., Olivier, P. (eds.) Diagrammatic Representation and Reasoning. Springer Verlag, Berlin (2002)
Shah, P., Mayer, R.E., Hegarty, M.: Graphs as aids to knowledge construction: signaling techniques for guiding the process of graph comprehension. J. Educ. Psychol. **1991**(4), 690–702 (1999)
Simkin, D., Hastie, R.: An information-processing analysis of graph perception. J. Am. Statist. Assoc. **82**(398), 454–465 (1987)

Sripada, S.G., Reiter, E., Hunter, J., Yu, J.: Segmenting time series for weather forecasting. In: Macintosh, A., Ellis, R., Coenen, F. (eds.) Proceedings of ES2002. pp. 193–206 (2002)
Tufte, E.R.: The Visual Display of Quantitative Information. Graphics Press, Cheshire, CT (1983)
Welford, A.T.: Attention, strategy, and reaction time. In: Kornblum, S., (ed.) Attention and Performance IV, pp. 37–54. Academic, New York (1973)
Wickens, C.D., Carswell, C.M.: The proximity compatibility principle: its psychological foundation and relevance to display design. Hum Factors **37**(3), 473–494 (1995)
Yu, J., Hunter, J., Reiter, E., Sripada, S.: Recognising visual patterns to communicate gas turbine time-series data. In: Macintosh, A., Ellis, R., Coenen, F. (eds.) Proceedings of ES2002. pp. 105–118 (2002)
Zacks, J., Tversky, B.: Bars and lines: a study of graphic communication. Mem Cogn **27**(6), 1073–1079 (1999)

Authors' Vitae

Dr. Stephanie Elzer Department of Computer Science, Millersville University, Millersville, PA 17551, USA.

Dr. Elzer is an Assistant Professor of Computer Science at Millersville University in Pennsylvania. She received her B.S. degree in Computer Science from Shippensburg University, and her M.S. and Ph.D. degrees in Computer Science from the University of Delaware. Dr. Elzer's research interests include natural language processing, cognitive modeling, user modeling, and human–computer interaction.

Dr. Nancy Green Department of Mathematical Sciences, University of North Carolina at Greensboro, 383 Bryan Building, Greensboro, NC 27402-6170, USA.

Dr. Green is Associate Professor of Computer Science at the University of North Carolina at Greensboro. Previously at Carnegie Mellon University, she worked as a Postdoctoral Research Fellow in the NL-SOAR research group, and as a Systems Scientist in the AutoBrief research group. Dr. Green received her M.A. degree in Linguistics from the University of North Carolina, her M.S. degree in Computer Science from the University of Pennsylvania, and her Ph.D. degree in Computer Science from the University of Delaware. Her current research interests include multimedia natural language generation and argumentation systems for healthcare applications.

Dr. Sandra Carberry Department of Computer Science, University of Delaware, Newark, DE 19716, USA.

Dr. Carberry is Professor of Computer Science at the University of Delaware. She received a B.A. degree in mathematics from Cornell University, a M.S. degree in computer science from Rice University, and a Ph.D. degree in computer science from the University of Delaware. Dr. Carberry pursues research in natural language processing, user modeling, and artificial intelligence, with an emphasis on interactive systems. She has served on many program committees and editorial boards, and is on the executive board of the Association for Computational Linguistics.

Dr. James E. Hoffman Department of Psychology, University of Delaware, 213 Wolf Hall, Newark, DE 19716, USA.

Dr. Hoffman is a Professor of Psychology at the University of Delaware. He received his B.S., M.A., and Ph.D degrees from the University of Illinois at Champaign-Urbana. Dr. Hoffman does research in the area of visual attention using eye-movement recording and event-related brain potentials. He is particularly interested in applying methods and findings from basic research in attention to various applied problems, including user interface design and computer-based training methods.