UNDERSTANDING THE STRUCTURE OF INFORMATION VISUALIZATION THROUGH VISUAL METAPHORS

by

Caroline Ziemkiewicz

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Information Technology

Charlotte

2010

Approved by:

____

Dr. Robert Kosara

____

Dr. William Ribarsky

____

Dr. Aidong Lu

____

Dr. Paula Goolkasian

____

Dr. Kent Curran

ABSTRACT

CAROLINE ZIEMKIEWICZ. Understanding the Structure of Information Visualization Through Visual Metaphors. (Under the direction of DR. ROBERT KOSARA)

Information visualization is an increasingly widespread way to present and analyze complex data, but there is much we still do not know about how people understand visually presented information. Every visualization contains certain assumptions about the structure of its information: how the data can be broken down into pieces, how those pieces relate to one another, what actions can and cannot be performed with the data, and so forth. Yet information visualization still lacks the language and theory to analyze these properties of visual information structure. I propose that these structural properties can be thought of as visual metaphors that drive a visualization, analogous to the verbal metaphors that structure abstract information in speech and writing. In this model, people analyze visual relationships among shapes and patterns in a visualization in the same way that they analyze other kinds of visual scenes, then metaphorically interpret those visual relationships as conceptual relationships. I have grounded this proposed model through empirical studies showing how metaphors affect visualization use and how minor structural changes can have significant effects on the way people interpret visual information. I argue that this framework sheds new light on the importance of design and conceptual structure in visualization and can substantially improve future techniques and evaluation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: UNDERSTANDING VISUAL STRUCTURE

Information visualization (infovis), the study of visual methods for presenting and analyzing data, is a young field just beginning to approach maturity. Systems such as Tableau [3] apply visualization research to business data analysis; social visualization sites like Many Eyes [58] have made infovis a tool for discussing topics from politics to health to movies; and news media are beginning to show interest in infovis as a way to communicate and analyze information, with newspapers such as *The New York Times* introducing sophisticated infographics based on infovis methods [23].

With this growth in the stature of the field has come a greater interest in the theory of how visualization works. Despite the many new techniques introduced by researchers over the past two decades, major gaps remain in our understanding of how people communicate and reason with visual information, limiting our ability to design and evaluate these techniques. Infovis theory has attempted to fill in these gaps, but has not consistently gone beyond low-level perceptual issues such as color use and the properties of visual marks. This leaves out a consideration of visual information structure, or how a visualization treats the parts of its data as relating to one another. A better understanding of how visualization works would mean understanding the structure of a visualization as well as its low-level perceptual properties.

Visualization is at a point in its development where its practitioners frequently find themselves grappling with big questions about the nature and purpose of their field. These include fundamental questions about how visualization works: how do people interpret visual forms as information? The answer we give to this question has far-reaching consequences for infovis practice. It determines what aspects of a visualization we consider essential or superficial, as well as what information we expect a visualization to express to a viewer. And these expectations in turn affect how we design, evaluate, and judge every visualization we make.

In many ways, however, the answers to this question have not evolved greatly since the early days of visualization. The classical view of visualization sees it as a process of encoding numerical or categorical values as visual (or *retinal*) variables like size, distance, or color, which are then decoded by the viewer to reconstruct the original information. This variable encoding model is the simplified essence of Bertin's *Semiology of Graphics* [9] and the years of visualization theory that have built upon it.

And yet, this account fails to cover much of what seems to happen in visualization. Shneiderman's classic visualization mantra [51], which states that the stages of infovis use are, "Overview first, zoom and filter, then details on demand," suggests the importance of a user's initial high-level view of the data in framing further analysis. Yet the variable encoding model focuses entirely on how low-level details are read, which leaves out key information that can be found in the overview. How does a user grasp the gist and high-level patterns in a visualization before exploration even begins?

The current model of how visualization works is a reductionist one: numbers become distances or sizes, which become numbers again when perceived by a user. Reductionist models are useful for breaking a complex phenomenon like visualization into manageable pieces, and the variable encoding model has indeed contributed much to visualization design and practice. But to explain how the overall structure of a visualization affects a viewer, we need a holistic model as well. The variable encoding model implicitly assumes that one shape—be it bar, circle, or something more exotic—is as good as another, provided the visual variable being used to encode the information can be perceived accurately. Likewise, it assumes that the perception of a variable is independent of its surrounding visual context.

However, there is already evidence that these assumptions do not entirely fit what happens when a viewer uses a visualization. Even in cases where data values are held constant, changing the design of a visualization or chart can have significant effects on how people see the data. A striking example comes from Elting et al. [22], who found that changing the presentation style of simple charts of clinical study data could lead to significant differences in the number of errors made by physicians. This case shows the potentially serious consequences of ignoring the effects of visual structure on the inferences and decisions people make when using infovis.

Another potential consequence of glossing over structural properties is illustrated by the case of inconsistent evaluations that originally inspired this research. Evaluation of the efficiency and usability of new visualization methods is necessary for application and progress in the field, but as Chen's meta-analysis of visualization evaluation papers [15] suggests, there is little agreement in the findings of the various usability studies performed by visualization researchers over the years. This is no doubt due

Table 1: Are variances in the metaphors of questions asked responsible for the inconsistent performance of a single visualization method across evaluation studies? We looked at several recent tree visualization evaluation papers that both included a treemap in their comparison and published their task questions. For each paper, we counted the number of questions or task descriptions that reflect a containment metaphor, i.e., those which used words like *contains*, *has*, *inside*, or *within* when describing relationships among nodes in the hierarchy. We then ranked each of the methods in the comparison by average response time overall, with 1 being the fastest method. Where exact response time was not reported, we estimated it based on results graphs. This informal meta-analysis suggests a possible relationship between metaphor compatibility and response time.

Paper	**Containment Questions**	**Treemap Ranking**	**Example Question (emphasis added)**
Andrews and Kasanicka [5]	6 of 8 (75%)	1 out of 4	"Find the deepest subdirectory *inside* the directory 'pad++."'
Kobsa [37]	6 of 15 (40%)	2 out of 6	"Find the directory that *contains* the most .png type files."
Stasko et al. [52]	4 of 12 (33%)	2 out of 2	"Identify a directory *containing* files of a particular type."
van Ham and van Wijk [57]	0 of 5 (0%)	3 out of 4	"Users had to indicate *level* of a predetermined node."
Barlow and Neville [7]	0 of 5 (0%)	4 out of 4	"Participants counted the number of *levels* in the tree."

in large part to the lack of standard experimental procedures and benchmarks, but as Chen suggests elsewhere [14], another factor is the lack of understanding of the cognitive processes at work in visualization use.

For example, hierarchy visualizations have been frequently subject to evaluation studies. Hierarchies are naturally applicable to a range of global and local information retrieval tasks, and many novel methods for visualizing them have been devised. However, there is little consensus across the existing evaluation papers in this domain, even when the same visualization methods and similar tasks are used. Most of these evaluations include a treemap [50], a method for displaying a hierarchy as a collection of rectangles nested within one another. While the treemap has long been considered a success story for infovis, these evaluation papers sometimes find the method to be the most efficient of those studied, sometimes the least efficient, and sometimes in between. Moreover, the studies often include highly similar task questions, leaving it nearly impossible to consistently interpret the treemap's strengths and weaknesses.

It is certainly possible that this inconsistency is entirely due to differences in the treemap implementation, the datasets used, the other methods studied, and the experimental designs. However, although the tasks used are often logically equivalent, we noticed a trend in the wording of the task questions (Table 1). Some of the researchers tended to word questions in terms of *levels* of a hierarchy; some tended to word them in terms of nodes in a hierarchy *containing* one another; other questions were worded in terms of *parent-child relationships* or *tree branches.* These differences in wording reflect different metaphors used to explain hierarchies. What we found in a survey of hierarchy visualization evaluation papers is that those researchers who used more *containment* metaphors seemed to find the treemap more efficient than those who used more *levels* metaphors.

This suggests that priming participants to think about hierarchies in terms of levels may influence their ability to understand the treemap, which uses an unusual spatial metaphor of containment (Figure 1). A focus on levels would seem to favor a traditional node-link diagram or an icicle plot, which indeed seem to perform more efficiently in the studies that reflect this focus.

If a visualization method can be made more or less efficient based by changing how a user thinks about the data, then a model of infovis as merely translating visual properties into data properties cannot capture the whole picture. There must be some

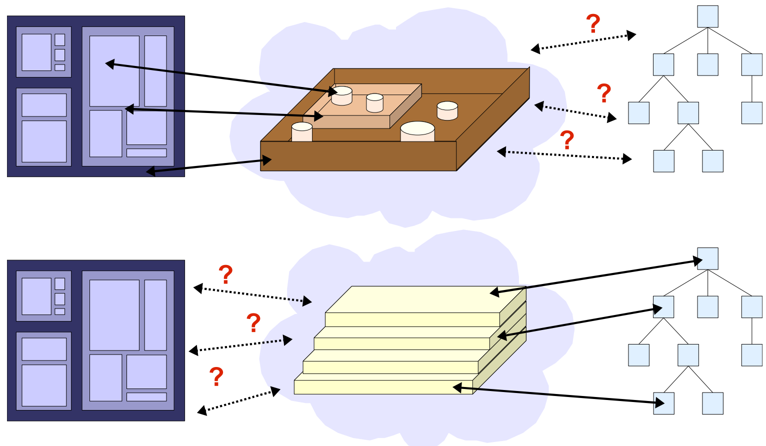


Figure 1: Understanding a visualization involves matching its visual structure to one's mental structures. If a user conceives of hierarchy data as a series of boxes nested inside each other, the correspondence to a treemap should be a natural fit. If she conceives of the hierarchy as a series of higher or lower levels, however, the correspondence will be more difficult and may take longer to process. For a visualization that employs a levels metaphor, like a node-link diagram, the reverse would be true.

interaction between how a user structures information and how a visualization structures information. Consistent evaluation in infovis may be difficult in part because we lack a framework for including this question of structure in our interpretation of results.

The case of tree visualization evaluation suggests metaphor as such a framework. Metaphors are known to structure information in language, they are well-studied in their verbal form, and many of these metaphors have a spatial or visual component that can be easily applied to visualization. It is also not uncommon for visualization designers and researchers to talk about infovis methods in terms of visual metaphors [19], suggesting that this is an intuitive way of thinking about infovis. Cognitive science research on verbal metaphors also suggests a number of predictions that can be made about how metaphors affect people, so this theory can be tested.

The purpose of this work is to establish the usefulness of visual metaphor as a framework for visual information structure and to analyze how metaphors influence infovis in practice. To establish this, this thesis will argue that metaphors are a rich and flexible model for information structure, that it is valid to treat visual metaphors as analogous to verbal metaphors, and that such metaphors have an affect on visualization use in practice. In the process, I will begin building a framework for visualization structure based on the findings and lessons learned from this research. While this framework is in its preliminary stages, this dissertation contributes a body of research that shows how structural properties can affect visualization use and perceptions of data, and makes an initial attempt to place these effects in the context of a theoretical framework driven by visual metaphors and perceived dynamics.

The first step in this process is to explore how information structure has been explored in visualization research to date, to identify what is missing and what visual metaphors can do to fill in those gaps.

CHAPTER 2: THEORIES OF VISUAL STRUCTURE

An information visualization, like any artifact used for communication and reasoning, is a representation system. This system includes correspondences between low-level properties of the data and the image, which is the information captured in the variable encoding model. However, it also includes a system for fitting those properties into a larger picture: the visual information structure. This structure provides context for individual data items, suggests patterns and relationships in the overall data, and assists the user in reasoning about visually presented information.

Current infovis theory has much more to say about the low-level data encoding side of this representation than about the high-level structural side. Theories in infovis and diagrammatic reasoning that do consider the importance of visual structure tend to be either fairly vague or to focus on spatial layout as another encoding dimension. However, there is work in the related field of human-computer interaction (HCI) that takes a more concrete view of how visual interfaces suggest structural properties of systems, which suggests a possible way forward for infovis theory in this regard. The importance of finding a way to integrate visual structure into infovis theory is shown by work in cognitive science that highlights the strong effects that structure and context can have on the perception of visual information. In this chapter, I will present and discuss this background in visual structure theories in terms of the attempt to make infovis theory more structurally sound.

2.1 Visual Structure in Infovis Theory

Infovis theory has most often adopted a model of visualization as information extraction. This model focuses on how data are transformed into visual encodings, and how a user then translates those visual encodings into internal knowledge. As a result, theory of this kind tends to be largely concerned with object-level rather than global properties. When structure is considered, it tends to be restricted to a question of what data attributes influence an object's position in space.

The seminal work in visualization theory is Bertin's *Semiology of Graphics* [9]. Although Bertin was at the time writing about static diagrams, his work has been highly influential in modern infovis. Bertin lays out a thorough system of information graphics, defining "marks" as the primitive graphical object whose visual and spatial properties are based on a mapping with underlying data. A mark can be any visual element, such as a shape, line, area, or point, that represents information. He refers to these visual properties as retinal properties, e.g., color, size, shape, and location. Based on psychological knowledge about perception, he then provides guidelines for the mapping of these properties to different types of data, such as categorical, ordinal, and numerical: color is best suited to categorical data, position is the most precise mapping for numerical values, and so forth.

Bertin also considers spatial structure in his work, primarily focusing on the image plane and how marks are positioned on it. He calls systems of planar organization "imposition" and sorts them broadly into diagrams, networks, maps, and symbols, which can be further classified by the coordinate system used. This part of his theory has been less broadly influential on infovis practice than the retinal properties, perhaps because it is less thorough and does not provide such clear guidelines. Another reason may be that the retinal properties were founded on scientific knowledge about the capabilities of the human visual system, and no equivalent knowledge existed at the time about how people understand visual structure. However, when visual structure has been considered in infovis theory, it has usually resembled Bertin's construction.

Similarly to Bertin, Cleveland and McGill's work on graphical perception [17] explains the comprehension of information graphics through elementary perceptual tasks, such as discerning angle, direction, area, and curvature of visual marks. Having identified these tasks, they describe common diagram types like bar charts, pie charts, and scatterplots in terms of which tasks are used to encode and decode data. Like Bertin, they go on to make recommendations on the suitability of certain graphics based on human perceptual abilities. Their theory is based on the idea that reading visual information is a process of extracting information by decoding the visual mapping.

These two works have together had a foundational influence on theoretical discussion of information visualization. In many cases, this influence is direct and explicit: for example, Mackinlay [42] employs Bertin's classifications of visual marks and Cleveland and McGill's recommendations in his system for automating graph design. Card and Mackinlay [13] also use a Bertin-inspired system to describe and classify visualization methods in a taxonomy. In their model, visualization methods are coded according to mappings between data variables and retinal variables; for example, data variables are first coded by data type (i.e., nominal, ordered, or quantitative) and then by what retinal or other visual property they are mapped to in a visualization. What is striking about this paper is that, when they apply this model to describing a number of actual infovis systems, it is almost always inadequate to the task. Nearly every encoding they present includes asterisks and question marks to indicate special cases, uncertainty about the visual variables being mapped, or what the authors call "non-semantic use of space-time." While this taxonomy makes a heroic attempt to unify data description and visualization description under a single model, the awkwardness of the fit seems to suggest that there are aspects of this visual mapping that do not easily fall under variable encoding.

In other cases, the influence is more subtle, and reflects the emphasis on marks and their visual properties in a broad range of ways. Wilkinson's grammar of graphics [61] attempts to define a language for combining these basic graphic elements. This grammar takes an object-oriented approach in order to define generalized designs of graphical representations of data. Like Bertin, Wilkinson considers structure only in terms of coordinate systems—that is, how the position of marks is determined. Shneiderman's task by data type taxonomy [51] classifies data by a similar set of structure types: one-dimensional, two-dimensional, three-dimensional, multidimensional, temporal, tree, and network. Although these classifications refer to inherent data properties, not visual structures, they are nonetheless influenced by assumptions about on-screen positioning, or there would be no reason to separate two- and three-dimensional data from the multidimensional category.

This influence is also present in taxonomies that classify visualization methods by how they encode data, such as Chi's data state reference model [16]. This model expands on the steps involved in translating data into visual form, then defines the behavior of a broad range of visualization methods at each step. This is similar to Card and Mackinlay's system, but is more process-oriented, emphasizing the encoding as a transformation rather than a simple translation. While it is useful to expand on what is meant by variable encoding, and what this process actually entails, it is still an expansion on a narrow definition of what is going on in the use of infovis.

A basic assumption of this area of theory, made explicit in Cleveland and McGill but implicit elsewhere, is that understanding a visualization is a process of information extraction. That is, there is some encoding from data property to visual property, and all a user does to gain knowledge from a visualization is reverse that encoding. This viewpoint sees all the activity of using infovis happening at the level of individual graphical marks; it does not allow for overall structural impressions having a significant impact on understanding.

There have been many practical benefits of this line of research, such as its application to automatic view generation in the visual analysis system Tableau [43]. Building on his previous, more theoretical work in automated graph design [42], Mackinlay provides users of Tableau with the option of automatically choosing the best graph for their data, based on the type of data dimensions being visualized. The variable encoding model has also provided a framework for usefully including knowledge of perception in visualization research. However, a body of theory that concerns only object properties is in danger of missing the forest for the trees. It is in some ways surprising that infovis has taken such a narrow view of visual information, since the closely related field of human-computer interaction (HCI) has dealt extensively with the idea that a visual interface represents structure.

2.2 Visual Structure in Human-Computer Interaction

In human-computer interaction (HCI), the idea that an interface (the system of input methods available when using a computer program) contains information about how its components fit together and how they can be used is a natural one. A common way of talking about this is in terms of a user's mental model of a system [48]. That is, when faced with a novel piece of software, a user tries to figure out how it works and what interactions are possible based on the appearance of interface components. These perceptions of form and function compose the user's mental model, which is used to make predictions about how to achieve a goal using the interface.

There is evidence that these mental models, far from being a purely abstract design concept, can have a powerful effect on memory and reasoning in interface use. Kieras and Bovair [35], in a series of experiments, presented participants with a novel device consisting of various switches and flashing lights, then taught them how to use the device either by rote (that is, by explaining what steps to take to achieve a specific result) or by giving them a model of the device's purpose and how it works, describing it in *Star Trek*-inspired terms as a control panel for a "phaser bank" and assigning purposes to the various interface components. Users given a meaningful model of how and why a device works were not only more able to remember memorized tasks using the device, but were also more likely to spontaneously find a more efficient way to perform the task.

This work shows how important structure is for understanding a complex system. While our purpose in infovis is not necessarily to solve problems (although it can be in some cases), the argument can still be made that exploring a dataset is a matter of learning a model for a complex system of information. Mental models are therefore a useful way to think about how a user comes to understand a dataset.

While mental models are a good way of thinking about how people conceive of the structure of software systems, the question of how people perceive that structure is perhaps a more pressing one. That is, how do people construct a mental model of a system, given the appearance and function of its interface? One of the most common ways to discuss this process in HCI is in terms of perceived affordances. The concept of affordances is originally derived from Gibson's ecological perception theory [29]. Gibson framed perception in terms of what actions a given animal sees its environment as affording. For example, a solid, flat surface affords supporting the animal's movement, while a smooth, sloped surface affords sliding downwards. In all cases, affordances are relative to the viewer; a given environment affords different actions to a mouse and to an elephant. Any animal faced with a given environment will automatically perceive such potentials for movement or action based on apparent physical properties and the animal's own abilities.

In HCI, the concept is used in a slightly different fashion, to refer to aspects of a visual interface that suggest potential actions to a user [47]. For example, an interface element that is styled to look like a physical toggle button suggests to the user that it can be pressed. The general model of visual structure in HCI, then, is that people view an interface in terms of its perceived physical affordances, derive predictions about what actions they can take based on those affordances, and then derive a mental model of the system by taking those actions and seeing how they meet their predictions.

Given the amount of research overlap between HCI and infovis, it is surprising that visualization is rarely thought of in terms of what mental models a technique suggests to a user. There are two likely reasons for this. The first is that HCI assumes that the systems it deals with are interactive, so the ability of a user to predict the outcome of her actions is an obvious consideration. Infovis, on the other hand, builds on a history of static depictions of data; interactivity is a more recent development for the field. Consequently, the ability of a user to perceive data accurately is the primary consideration.

The other reason is the lack of well-defined tasks in infovis. Having a model of how a system works is obviously necessary if you need to use it in pursuit of a goal. Knowing what you can do and how to do it are prerequisites for solving a problem. But in infovis, we don't necessarily know what problem we're trying to solve. The tasks we feel visualization systems are meant to address are vague ones like understanding a dataset, forming hypotheses, pattern recognition, and exploration. These are important tasks, and the possibility of systems that can help perform them is what excites people about visualization. But they are also tasks that lack a clear end state. Perhaps this aspect of visualization makes structural properties of the interface seem less important than in other domains.

However, even a task without a clear goal can benefit from structure, even if the contribution of a user's mental model seems less direct or obvious. Some of the ways that visual structure can affect understanding and general reasoning are illuminated by work in diagrammatic reasoning and visual cognition.

2.3 Visual Cognition of Diagrams

While the information processing approach has provided a way to apply perception research to information visualization, it is less well-suited to understanding visualization from the perspective of higher-level cognition; that is, not only how people perceive information, but how they learn, reason with, and remember information. This cognitive perspective forces us to consider the structural properties of visualization and how they affect not only what information is extracted but how that information is understood.

Theories that focus on reasoning with visual representations include Stenning and Oberlander's view of diagrams and language as logically equivalent yet supporting different facilities of inference [54]. That is, by making certain aspects of a problem specific through visual representation, diagrams such as Euler circles can make certain problem constraints explicit and therefore restrict potential inferences to a smaller, valid subset. Similarly, Larkin and Simon [41] consider the differences between graphical and verbal representations as differences in what information is made salient and explicit. In a graphical representation, information is naturally organized by location, while in a sentential representation it is organized sequentially. This makes graphs more useful for, e.g., solving geometry problems, and language more useful for problems that require logical reasoning. The authors consider what effects the structure of a representation has on understanding, although they focus on the very broad differences between words and pictures rather than defining differences among types of graphical structure.

The importance of such differences, however, is illuminated by the extensive body of work by Tversky and colleagues on how people interpret information presented in different visual representations. For example, the authors presented the same simple two-point data as either a bar chart or a line graph and asked for users' interpretations [63]. They find that those viewing a bar chart tended to describe the diagram as depicting two separate groups, while those viewing a line graph described the data as a trend. This effect holds even when the interpretations conflicted with the labels on the data points. For example, a line graph showing the average height of males versus females prompted one participant to describe the chart as saying "The more male a person is, the taller he/she is." These findings and others are further discussed as examples of how schematic figures such as bars and lines are interpreted in varying contexts [56]. Many of these figures have seemingly natural interpretations; for example, lines between marks imply a relationship between the represented objects, while contours are used for grouping objects. However, in many cases context aids the interpretation of ambiguous primitive features such as blobs and lines by fitting their relevant properties to task demands. Understanding the cognitive basis for these primitive features and how they can be altered in context would go a long way towards explaining how visualization works.

This work has a particularly direct application to infovis, but it also recalls a broader area of visual cognition that looks at how people use diagrams as an external representation to aid in reasoning. Gattis and Holyoak [28] argue that the power of graphical representations go beyond Larkin and Simon's view that they merely allow for more efficient information access in certain cases. Rather, they see diagrams and graphs as having a special role in supporting reasoning by mapping conceptual relationships to spatial ones, so that inferences about spatial properties can be extended to inferences about the represented information. This view is supported by a number of studies on diagrammatic reasoning, such as Bauer and Johnson-Laird's finding [8] that diagrams improve reasoning if they visually represent meaningful constraints in a problem and Glenberg and Langston's demonstration [30] that diagrams only improve efficiency when their spatial mapping is conceptually meaningful. This work taken together suggests that graphics can assist in problem solving, but only when their spatial structure is meaningful in some way. The question of what structures are meaningful and which are not, however, is not easily answered by existing work.

While this work suggests the importance of structure to the understanding of information visualization, it offers no clear framework for discussing and analyzing that structure. While they intuitively seem to be talking about the same thing, researchers from different fields and perspectives may refer to these structural properties as visual framing, spatial layout, graph types, and so on. A common language and theory for discussing the effects of structure is necessary to integrate it into visualization practice, as Bertin's conception of retinal properties has provided a common language to deal with object properties. A promising source for this theory is visual metaphor.

CHAPTER 3: METAPHORS AS INFORMATION STRUCTURE

Traditionally, metaphors are figures of speech in which the properties of one object or concept are mapped to the properties of another. This mapping can reveal some underlying similarity between the two things or cause one or both of them to be seen from a new perspective. While the classical idea of a metaphor is a relatively explicit comparison used for poetic or rhetorical purposes (e.g., "All the world's a stage"), there is considerable evidence that everyday language uses less obvious metaphors to structure abstract concepts. This conceptual structuring is often based on metaphors of space and vision, making it potentially invaluable for understanding how visualization organizes information.

3.1 Verbal Metaphors

There has been extensive research on how verbal metaphors shape our understanding of information conveyed in language. This work was pioneered by Lakoff and Johnson [40], who argue that, far from being an unusual poetic device, metaphor is used constantly in everyday speech. Rather than being used for dramatic effect, these metaphors are mostly subconscious and play an important role in giving recognizable structure to abstract concepts like time, emotion, and ideas.

In Lakoff and Johnson's formalization, metaphor is a mapping of structural properties between a source domain and a target domain: e.g., *time is a landscape we move* *through*. *Landscape*, the source, provides assumptions and concrete properties to the more abstract target *time*, allowing us to say that we *move through* time or that an upcoming event is *ahead of us*. These metaphors can also form overarching systems. For example, the "conduit metaphor" is a metaphorical system in which *communication is transfer of objects*, and is made up of many common lower-level metaphors such as *ideas are objects*, *ideas are contained in language*, and *ideas are projectiles which the sender conveys in some manner*. Such complex conceptual structures recall the mental models discussed in Section 2.2.

An interesting aspect of this metaphor theory is that, by mapping abstract ideas to physical objects, they often lead people to make inferences about those ideas based on physical dynamics [39]. For example, the metaphor *anger is heat*, reflected in phrases such as "hotheaded" and "to make one's blood boil," may motivate folk psychology inferences such as "bottling up anger will cause you to explode." This suggests that the purpose of metaphor may be to make it easier to reason about abstract concepts by mapping them to more natural reasoning about the physical world.

Other researchers have worked to further describe and classify different types of metaphor. The distinction between poetic and everyday metaphor, for example, is one of conventionality; some metaphors are so commonly used and a conventional part of language that we don't notice them, whereas more novel ones may require us to actively figure them out. Bowdle and Gentner [11] have found evidence that the two kinds of metaphor are processed differently. Conventional metaphors are comprehended as quickly as literal sentences, while novel metaphors are slower and require an active comparison between the source and the target. While novel metaphors may take longer to process, they may also be more useful in reasoning with difficult information, as evidenced by findings by Cooke and Bartha [18] showing that graduate and postgraduate psychology students use more novel metaphors for psychological topics than undergraduate students.

Another important distinction between types of metaphor deals with the complexity of the mapping between source and target. Lakoff and Johnson [40] classify metaphors as either ontological, orientational, or structural. Most of the metaphors I have discussed up to this point are structural metaphors, in which the source and target are both complex domains and the metaphor involves extensive mapping between features in each. *Time is a landscape* and *anger is heat* are good examples of structural metaphors.

Ontological metaphors are far simpler mappings, in which the target concept is mapped to a broad category such as "object" or "substance." For example, *ideas are objects* is an ontological metaphor, as is *emotional states are containers* (e.g, "I'm *in* love," "He's finally *out* of his bad mood"). These metaphors are so basic as to be nearly invisible, but they play a vital role in structuring thought. Thinking of ideas as relationships rather than objects, for example, would fundamentally alter our basic metaphor for communication. Instead of communication being the transfer of objects, it would be seen as the building of connections—a counterintuitive but not inconceivable thought. Almost as fundamental are orientational metaphors, which map a concept or quality to a direction or other simple spatial property. Examples include the nearly universal *more is up*, as well as *power is height* and *importance is centrality.*

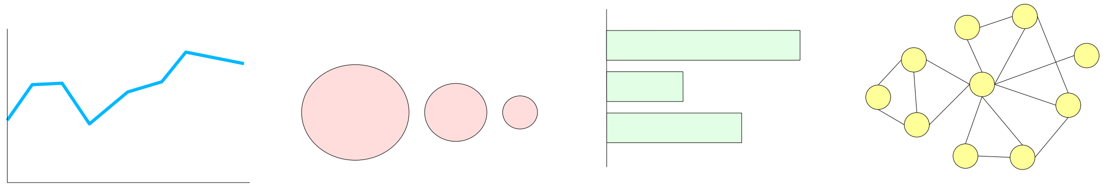


Figure 2: Many common verbal metaphors that use spatial concepts as source domains can be found in the most basic structural elements of infovis. Amount is frequently depicted as verticality (i.e., *more is up*), size, or length. Network graphs often place more important elements near the center; in a graph visualization, this usually means the most highly connected elements.

What makes these verbal metaphors potentially useful to visualization theory is that many of them are ultimately spatial or visual in nature. Infovis in general can be said to be based on the metaphor system *properties are physical properties*. This broad system includes any case where abstract target values are mapped to physical attributes, including such mappings common to infovis as *amount is size*, *amount is verticality*, *amount is length*, *importance is size*, and *importance is centrality* (Figure 2). While these examples suggest rightly that most visual metaphors used in visualization are orientational, there are also visualizations that reflect higher-level structural metaphors. ThemeRiver [32] clearly embodies the metaphor *flow of events is flow of water*, and visual analytics systems with a mind-mapping style, such as Jigsaw [53] and the Sandbox from Oculus [62], draw on the metaphor *facts are points (set up in spatial configuration).* The IN-SPIRE Galaxy View [33] draws on this metaphor as well as *ideas are light sources.*

Metaphors have proven to be a useful way to analyze the sometimes subtle effects of conceptual structure on the understanding of information in language. Given an analogous definition of visual metaphor, it is possible that similar analysis could be applied to understanding information visualization. As discussed, many verbal metaphors with a spatial component can be seen as directly applicable to visualization. Understanding what aspects of verbal metaphor are relevant to visual metaphor is a crucial next step.

3.2 Visual Metaphors

Visual metaphors are less well-studied than verbal metaphors, but they are hardly novel. They have been considered in the context of human-computer interaction (HCI), particularly in the communication of mental models to understand complex software. Additionally, the ability of visual metaphors to persuade through association has been studied in advertising research. This work provides an important bridge between conceptual structures in language and similar structures in visually presented information.

A traditional discussion of visual metaphor as it relates to visualization comes from Cox [19], who sees metaphor as a framework for discussing the assumptions and aesthetics of choices in visualization design. Cox introduces many of the theoretical concepts underlying this work, although her focus is ultimately on the high level design implications of visual metaphor. In the same vein as Cox's work, visualization researchers frequently discuss visual metaphors in terms of providing a design inspiration. For example, Hetzler et al. [33] discuss using a landscape metaphor to help users understand relationships among documents in a collection, while Weber et al. [60] refer to the spiral used in their time-series visualization as a "visualization metaphor."

As with verbal metaphors, some of the visual metaphors used in visualization are more poetic and novel than others; both Horn et al. [34] and Agutter et al. [4] have developed complicated glyphs to depict a patient's medical data, based on a human body and a heart surrounded by blood vessels respectively. Both glyphs are strongly suggestive of their source domain, which the authors argue helps to contextualize and structure the critical information they contain. Horn et al. further argue that more uniform visual depictions, such as radial plots, are more difficult to learn due to their lack of organizing metaphors.

This design-centered use of metaphors owes a debt to HCI, where metaphors have traditionally been seen as useful in helping users to develop mental models of complex software [10]. For example, the desktop metaphor of the Mac and Windows operating systems allows users to apply real-world knowledge of files and folders in an office to the management of digital information. The idea that metaphors assist in forming mental models is very close to our idea that they structure information, as discussed in Section 2.2. The primary difference is that the HCI view emphasizes the mechanics of software, while in visualization we are primarily interested in gaining knowledge and information—the "what" rather than the "how."

Another practical use of visual metaphors is in advertising, where they are a significant part of the visual rhetoric used to suggest views of a product. These are often akin to the more "poetic" metaphors in language, meant to be consciously understood and indeed to startle the viewer with their novelty. For example, an ad for dish detergent may depict miniature bulldozers cleaning grease off a dish. McQuarrie's studies of consumer response to visual metaphors [44, 45] show that not only do viewers tend to rate ads with visual metaphors more positively than those with more straightforward product pitches, but also tend to elaborate upon them. That is, viewers of an ad that employs visual metaphors are more likely to spontaneously produce positive inferences about the product. Naturally this is good news for the advertisers, but it also has implications for our own understanding of visual metaphors. Does this kind of elaboration also occur with the more subtle and abstract metaphors of visualization, and if so, is that a good thing?

The effect of visual metaphors in mental models and consumer persuasion clearly relates to findings in the visual cognition of diagrams (Section 2.3). Mental models, elaborations, and graphical context are all ways of discussing how visual structure affects information. If we take visual metaphors as a framework for visual structure, we may be able to synthesize much of this disparate knowledge. This would allow for a more consistent and deeper understanding of visual structure and how it shapes information.

CHAPTER 4: A VISUAL METAPHOR MODEL OF INFOVIS

Having argued for the importance of visual information structure, the next step is to outline what a useful structural theory would look like, and what is needed to establish such a theory. Given the evidence from diagrammatic reasoning that structural differences can have a powerful effect on reasoning and inference-making, an important part of this theory would be a way to distinguish meaningfully between visual structures. Currently we have no standardized language for describing the structural differences between, say, a treemap and a node-link diagram, although such differences are readily apparent.

Our theory should also characterize how the process of visualization works beyond simple information encoding and decoding. Current models of infovis give us no way to predict what kinds of inferences are likely to be supported by a given visual structure, or how a user might read the patterns and relationships in a visualization, as opposed to individual data values. They also shed little light on how cognitive and environmental factors might affect the way a visualization is understood, assuming that a visualization expresses the same information independent of context.

I argue that visual metaphor makes a good candidate for a model of infovis that can fill in these theoretical gaps. One of the great advantages of metaphor as a model for information structure is that it is well-studied and can build on a large amount of existing literature. This existing work can provide useful taxonomies and descriptive language, as in the distinctions among ontological, orientational, and structural metaphors discussed in Section 3.1. Metaphors are also by nature very adaptable. While linguists and cognitive scientists have identified a number of common verbal metaphors, the number of possible metaphors is infinite; nearly any source and target can be combined into a metaphorical mapping. This generality means that even very idiosyncratic visualization techniques could be described as metaphors, giving us a common ground to compare traditional and novel techniques.

Finally, using visual metaphors as a model for visual information structure is desirable simply because it is so intuitive. As discussed in Section 3.2, visualization designers often speak of the visual metaphors in their work, and visual metaphors are also a common theme in interface design. Since we tend to think in this fashion anyway, it only makes sense to use this intuitive concept as a jumping-off point for a more detailed and rigorous model of how people interpret visual structure.

The general outline of this model is as follows. Faced with an information visualization, a viewer first sees only a collection of shapes. The viewer does read object-level data, but probably only for one mark or a very small number of marks at a time. More importantly, the viewer will be trying to organize the information into some kind of structure. This will involve some simple grouping based on proximity, connectedness, or similarity. But it also involves a higher-level interpretation of apparent relationships and dynamics between shapes and groups of shapes. These relationships are first seen as purely visual patterns, but those patterns are then mapped to conceptual relationships based on some visual metaphor, which may either be suggested directly by the visualization, taught to the viewer through instructions, or invented by the viewer in the absence of other information. This metaphor will then be used to understand individual data points in context.

There are a number of interesting implications to this model. For one, it supposes a much more active process of reading a visualization than the visual encoding model does. Rather than passively decoding data values, the user is continuously trying to fit information to an internal structure that more or less fits the structure of the visualization. This is, of course, very similar to the human-computer interaction model that sees users as fitting predictions and goals to a mental model of a system. If this is the case, then a user's ability to read data values should be affected by the goodness of fit between her internal information structures and the structure of the visualization, as performance with other kinds of interfaces is affected by the accuracy of the user's mental model [35].

Another consequence of this model is that we should expect significant differences in how people perceive the same data visualized using different visual structures. These differences should primarily affect perceptions of relationships among data items and higher-level patterns in the data as a whole. Drawing on the verbal metaphor theory which argues that metaphors support inferences about abstract concepts by analogy to predictions about physical dynamics, I propose that these perceptions of relationships will largely be driven by perceptions of physical dynamics among shapes in the visualization.

The purpose of this work is to ground the visual metaphor model of visual information structure by testing these implications in practice. In the chapters that follow, I will present empirical evidence that conflicts between internal and external information structure can affect visualization use, that even minor structural differences can affect data perception, and that simple visual dynamics may underlie the ability of structural elements to suggest conceptual relationships. This evidence argues not only that understanding information structure is vitally important to understanding how visualization works, but also that visual metaphors can contribute powerfully to this understanding.

CHAPTER 5: STRUCTURAL CONFLICTS IN INFOVIS USE

One of the notable aspects of the visual metaphor model of infovis is that it sees visualization as a much more active process on the user's part than the visual encoding model. In the visual metaphor model, understanding visualized information requires constant translation between internal and external information structures, and individual data items are always understood in the context of broader structural properties.

This process sheds light on our earlier finding that evaluations of tree visualizations seem to favor visualizations that embody similar metaphors to those used in the task questions (Chapter 1). By using different information structures in a task question and the visualization used to answer the question, the researchers may have introduced an extra structural translation step that confused or slowed down their participants.

If this is true, it suggests that ignoring structural properties is not just a matter of theoretical hair-splitting, but can have serious practical effects on an important part of visualization research. If our current model of how visualization works is leading to inconsistent evaluation, we will not be able to move forward in developing and validating new techniques until that model is revised. Additionally, if the effect of verbal and visual metaphor conflict is reliable, then the two types of metaphor must share an underlying similarity, motivating the use of metaphor as a way to describe visual structure. Therefore, the first step of this research is to test whether such structural conflicts do have a systematic effect on infovis evaluation.

5.1 Exploring Structural Conflicts Between Visual and Verbal Metaphors

The specific questions we chose to study in the first phase of this research [64] are as follows:

1. Are visual metaphors analogous to verbal metaphors?

2. Does priming a user with a particular verbal metaphor affect her ability to process an analogous visual metaphor?

Since verbal metaphors are known to influence how information is processed, we can use verbal instructions to prime a user to think in terms of a particular metaphor. We can then test whether a visual metaphor influences thought in the same way by testing whether that priming affects the speed at which visual information is understood.

Inspired by the effects seen in tree visualization evaluation studies, our hypothesis is that participants will be slower in responding to questions that reflect a verbal metaphor which is incompatible with the visual metaphor of the visualization they are using and faster in responding to questions that reflect a compatible verbal metaphor.

5.1.1 Performing Studies with Mechanical Turk

In this study and subsequent ones, we used Amazon's Mechanical Turk web service [1] to recruit and process participants. Mechanical Turk is an online marketplace provided by Amazon to facilitate the assignment of brief, simple tasks to a broad base of internet users. Users of the system are either requesters, who provide a task, or workers, who perform the tasks. Requesters post a listing on the site that describes their task, which is usually something that can be performed quickly and easily through a web browser: for example, tagging an image or evaluating a website design. The requester decides ahead of time on a payment for the task (usually less than a dollar) and the number of assignments. Workers browse these listings and choose a task based on its description. After the worker performs the task, she submits the result, and the requester can choose to either accept or reject the work unit. Additionally, the requester can choose to reward a bonus for especially good work.

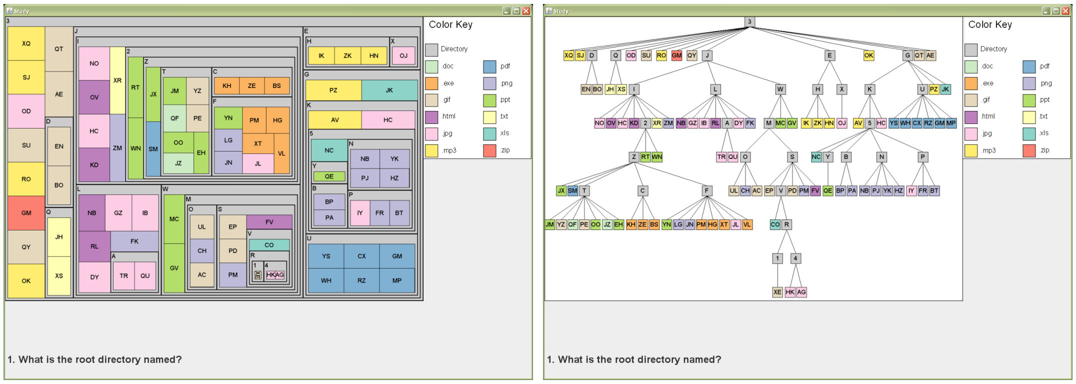
The nature of this service makes it well-suited to performing online studies of the kind presented in this dissertation [38]. Amazon provides tools for easily setting up simple surveys for users to answer, but it also makes it possible to embed a Java applet directly in their interface. This makes it possible to set up studies with more complicated functions and inputs, such as those described in later sections of this work. Amazon automates many of the details of the study process, including prevention of duplicate participation, payment, and advertising of the study. It also provides a number of advantages over a traditional university study. The demographics of the participant pool are much broader than in studies that exclusively use university students. Aggregating three studies over Mechanical Turk with a total of 366 participants, we found that the self-reported gender of our subjects was 57.4% female and 42.6% male, with age ranging from 18 to 64 and an average age of 32 (S.D. = 9.6). This service also makes it possible to run a large number of subjects very quickly; while participation rates vary throughout the day and week, in one case we finished a study with 50 participants in under two hours.

There are also some caveats to using Mechanical Turk. Other researchers have encountered widespread attempts to cheat the service [36], with participants entering nonsense responses or skipping questions in order to submit the task quickly. While these can be found and rejected upon review of the submitted tasks, it is more efficient to try and prevent or discourage such cheating in the study design itself. General disadvantages of online experiments also apply to Mechanical Turk studies, including a lack of ecological validity (that is, not knowing what other factors may be present in the participant's environment) and the fact that self-reported demographics data such as age and gender must be taken with a grain of salt. Overall, however, we found that the advantages of using Mechanical Turk outweighed the disadvantages.

5.1.2 Procedure

To study the effects of verbal metaphor and visual metaphor on task performance, we performed a study with 119 participants. The group included 64 females and 55 males, with age ranging from 18 to 57 ($M = 30.5$). 33 were students at the University of North Carolina at Charlotte, and 86 performed the study online through Amazon's Mechanical Turk.

During the study, participants were shown three hierarchical datasets in one of two tree visualizations: a treemap (Figure 3a) or a node-link diagram (Figure 3b). The visualization type varied between subjects, so that each participant saw only one type of visualization. The visualizations were created using the Infovis Toolkit [24], and we attempted to keep as many of the surface visual qualities constant across the two



(a) Treemap condition

(b) Node-link condition

Figure 3: These screenshots show our study design as it was seen by participants. Each participant viewed one of two types of tree visualization in three sessions, each of which showed a different but similarly complex dataset based on a hypothetical file structure.

representations as possible. The color scheme, window size, and label appearance did not differ between the two. Additionally, node size in the treemap was not given a meaning, to remove the possibility of questions that could be answered with one visualization but not the other. In order to focus entirely on the effect of visual representation, we did not include interactivity in either visualization.

The three datasets were described to the participants as representing hypothetical file hierarchies. Color was used to indicate file type, and the same color key was provided to the two groups. Files were named with random strings of two characters, and directories were named with a single random character or digit. These random names were meant to remove any possibility of the user answering the questions through inference rather than by consulting the visualization, as may be possible when using meaningful file and directory names.

Participants were initially told that the purpose of the study was to evaluate dif-

Table 2: The eight task questions given to participants in our study, framed in either a metaphor of hierarchy as containment or hierarchy as levels, as asked during the first trial session. For the subsequent sessions, the specific files and directories mentioned in the questions were altered to match the dataset being visualized, but the wording remained unchanged.

Containment Metaphor	**Levels Metaphor**
1. How many directories enclose the deepest file?	1. How many directories are above the lowest-level file?
2. How many total subdirectories are within the directory “S”?	2. How many total subdirectories are under the directory “S”?
3. How many files are immediately inside the directory “I”?	3.How many files are immediately under the directory “I”?
4. What is the deepest directory that contains both “XE.gif” and “KH.exe”?	4. What is the lowest-level directory that both “XE.gif” and “KH.exe” fall under?
5. Which directory immediately contains the most files of type “.pdf”?	5. Which directory can the most files of type “.pdf” be found immediately under?
6. What is the directory that immediately contains the directory “V”?	6. What is the directory immediately above the directory “V”?
7. Which directory contains the largest number of immediate subdirectories?	7. Which directory has the largest number of subdirectories immediately below it?
8. Which directory contains a deeper hierarchy: “G” or “M”?	8. Which directory has more levels under it: “G” or “M”?

ferent types of hierarchy visualization. After an initial training period in which they answered four questions and were given a chance to try again if they answered incorrectly, participants were asked eight questions about each dataset. Each question was answerable with a single keystroke: either a digit or one of the single-character directory names. During the experimental phase, participants were not informed if they answered incorrectly.

For each of the eight questions, we prepared two versions: one that reflected a “containment” metaphor, and one that reflected a “levels” metaphor (Table 2). The containment metaphor was considered to be more compatible with the treemap view, and the levels metaphor was considered to be more compatible with the node-link view.

Verbal metaphor was varied within subjects, in order to study the compatibility effect independently of individual differences in accuracy and response time. During their time with each of the three datasets, a participant saw four questions of the containment type and four questions of the levels type. The set of questions used for each dataset was counterbalanced from subject to subject, and question order was randomized. The result is that each participant, during each session, would answer a series of eight questions that randomly switched between a compatible and an incompatible metaphor relative to the visualization she was using.

For each question, we measured the participant's response time and whether they answered the question correctly. Altogether, participants answered twenty-four task questions. After the three sessions were complete, users filled out a short usability survey and were asked to write down any comments about the visualization they had used.

5.1.3 Results and Discussion

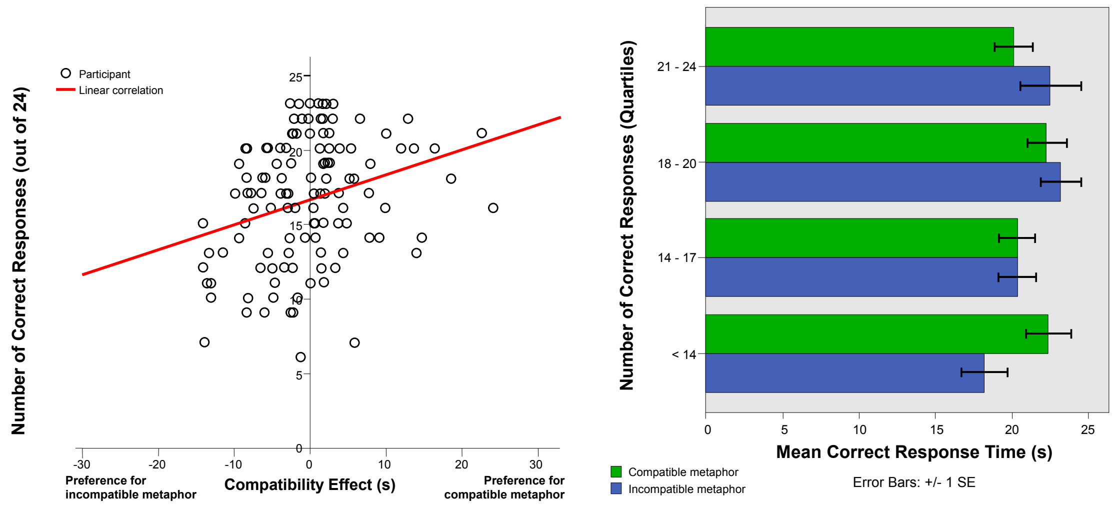
The results of our study, measured in response time and accuracy, suggest a complex picture of how metaphors affect visualization use. Participants slightly tend to show slower response time when responding to task questions with metaphors incompatible with the visualization they are using, although this effect was not statistically significant and is affected by a great deal of variance among participants. More surprisingly, we found that the amount by which any given participant performed faster on compatible questions was strongly correlated with that participant's overall accuracy.

To test the hypothesis that participants will perform faster on questions compatible with their visualization than on those which are incompatible, we first computed the participants' overall mean response time on incompatible task questions and on compatible task questions, with only correct responses considered. Contrary to our hypothesis, we did not find a reliable difference between compatible and incompatible response times overall.

However, there is a large amount of individual variance in performance among the participants. Some participants' response times favored the compatible metaphors by a very large amount on average, and some mostly favored the incompatible metaphor. To our surprise, this difference is not predicted by the type of visualization they were using. There were also no major differences in compatibility effect that might arise from the questions themselves. While some questions took longer to answer overall than others, there was no evidence that any question was more difficult in one metaphor or another.

One factor which did predict whether a given participant performed faster on metaphorically compatible questions, however, is overall accuracy. The average difference between a participant's incompatible response time and compatible response time highly correlates with that participant's number of correct responses across all three sessions, $r(117) = 0.29$, $p < 0.01$. That is, the degree to which a participant favors the compatible metaphor correlates with that participant's accuracy in using the visualization.

Furthermore, when controlling for the number of correct responses using a repeated measures analysis of covariance (ANCOVA), we did find a significant effect of com-



(a) Linear Correlation

(b) Participant Accuracy Groups

Figure 4: (a) We calculate each participant's preference for compatible metaphors by subtracting their average response time on compatible questions from their average response time on incompatible questions. High positive values therefore indicate a strong preference for compatible metaphors, while negative values indicate a preference for incompatible metaphors. When participants are plotted in terms of their preference and overall accuracy, there is a strong correlation between preference for compatible metaphors and accuracy. (b) Here we divide the participants into quartiles based on the number of questions they answered correctly (out of 24). Participants in higher-accuracy groups have a much higher tendency to perform faster on metaphorically compatible questions.

patibility on a participant's response time, $F(1,\ 117) = 11.07$, $p < 0.01$. Taken together, these findings suggest a close relationship between a user's understanding of a visualization and her ability or inclination to internalize its visual metaphors.

The findings of this study strongly support the need for better consideration and understanding of how structure influences the processing of information visualization. The fact that structural conflicts are correlated with user accuracy suggests that internalizing the visual metaphor is an important step in using any given visualization. Still, the fact that conflicts were only present for some users raises a number of questions which motivated the next phase of this research. The ability to predict which users will experience such conflicts is not only important for designing and interpreting user studies, but also for understanding what factors influence the perception and use of visual information structure.

5.2 Individual Differences in Understanding Visual Metaphors

While the results of our original study suggest that structural conflicts can have a significant effect on visualization performance, they also show that these conflicts are by no means insurmountable. This may mean that certain users have an easier time switching between two conflicting metaphors than others. Knowing what individual factors facilitate such metaphor switching would tell us a great deal about how people use visual information structure. To test the role of individual differences in internalizing visual metaphors, we conducted a study [65] to analyze potentially important factors such as spatial ability, personality, and metaphor preference in producing an interaction between visual and verbal metaphors in tree visualization use.

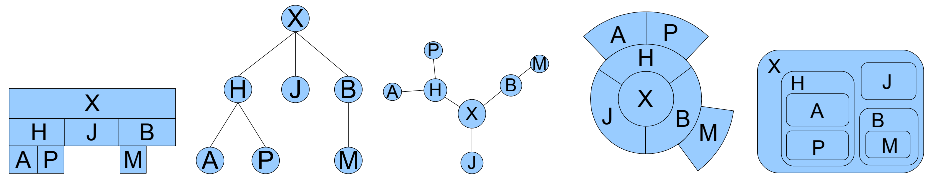
The design of this study is generally similar to that described in Section 5.1, although a number of improvements were made based on lessons learned in the previous study. In that design, visualization was varied between subjects and verbal metaphors alternated between compatible and incompatible in the same block. We believe the task-switching costs associated with this design may have limited our ability to clearly analyze the compatibility effect. Furthermore, we found great variation in the difficulty of the eight task questions used in that study across the two visualization conditions, and therefore limited the task questions in this study to the four tasks that appeared to be close in difficulty for both treemap and nodelink users. Finally, and most importantly, we added a number of pre-experiment questionnaires to analyze the individual factors that might affect a user's susceptibility to structural conflicts.

The hypothesis was that participants would be faster and more accurate on the compatible blocks, and that this effect would be stronger for more difficult tasks. The strength of structural conflicts would be influenced by the Openness dimension of a participant's personality and their spatial ability, since users with high scores on these cognitive factors may be able to more quickly adopt a novel visual metaphor. We further hypothesize that participants who showed a strong self-reported preference for one visual or verbal metaphor of hierarchies over the others will have lower accuracy overall and be more likely to experience interference from metaphors that conflict with their own.

5.2.1 Procedure

63 participants were recruited from Amazon's Mechanical Turk web service [1]. The participants were paid a base rate of $0.50 for their participation (which took about an hour), and could receive additional bonuses for answering questions correctly, for a total payment of up to $2.50. Participants were required to have 20/20 full color vision and be able to read and write in English. Of the participants, 40 were female and 23 were male. Age ranged from 18 to 54 ($M = 30.6$).

Participants initially filled out three scales in web forms meant to measure individual differences that may affect their performance in the study. Personality differ-



(a) Icicle plot (b) Node-link diagram (c) Radial tree (d) Tree rings (e) Treemap

Figure 5: During our pre-experiment questionnaires, we asked participants to rate five visual metaphors for hierarchies, based on common visualizations of hierarchy data. Participants were given a description of a simple tree structure and asked to rank these images from one to five in terms of how well they depicted the structure.

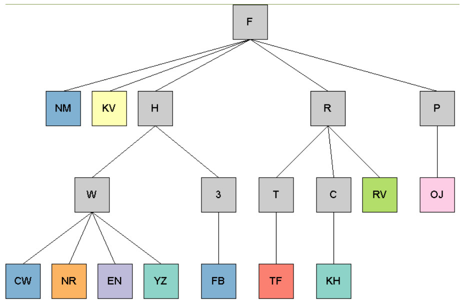
ences were assessed using the Mini-IPIP Big Five personality scale [20]. This twenty-question scale rates participants on five major personality dimensions: Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. We were primarily interested in the dimension of openness (or imagination) which measures a person's comfort with abstract and imaginative thinking, since we hypothesized this might predict a user's ability to switch between conflicting thinking styles.

Spatial ability was measured using the Form Board test (VZ-1) from the College Board *Kit of Factor-Referenced Cognitive Tests* [21]. This is a test for the cognitive factor known coincidentally as *visualization*, defined as "the ability to manipulate or transform the image of spatial patterns into other arrangements." We will subsequently refer to this factor generically as "spatial ability" to avoid confusion. In this test, users are given a target shape and a group of smaller shapes, and asked to find a combination of the smaller shapes that can be combined to form the target. We chose this test because the factor matches our view of visualization as a process of translating between spatial structures. Although we used an abbreviated version of the test for time reasons, the scores of our participant group ($M = 128.6, S.D. = 42.4$) were quite close to the baseline scores reported by the College Board ($M = 124.8, S.D. = 38.3$).

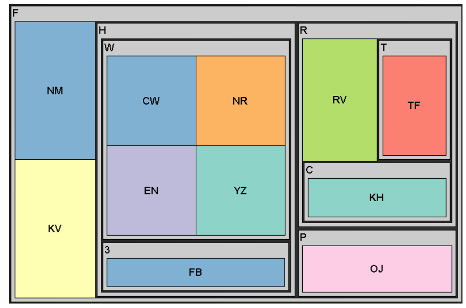
We also developed a simple scale to measure a user's preference for *levels* or *containers* metaphors in verbal descriptions of hierarchical relationships. Participants were given a description of a simple hierarchy, described as the department structure of a university. This description avoided as much as possible using strong metaphorical language in explaining the relationships between departments (e.g., "At the College of Humanities, there are two subdepartments: Art and Psychology.") Participants were then given a list of twelve statements about the university that were worded in either a *levels* or a *containers* metaphor, and asked to rate how well the statements described the university's department structure on a scale from one (Very bad description) to five (Very good description). For example, "The Marketing department is inside the College of Business" versus "The Marketing department falls under the College of Business." Finally, we asked participants to rank their preference for five different visual metaphors for this same hierarchical structure based on common visualization methods (Figure 5).

During the test portion, participants answered simple questions about a hierarchy (described as the files on a computer hard drive) visualized as either a treemap or a node-link diagram (Figure 6). These visualizations were similar to those used in the original study, but we included smaller datasets in order to test whether difficulty had an effect on structural conflicts. We generated four hierarchical datasets to visualize, two of which were small four-level hierarchies and two of which were more complex eight-level hierarchies.

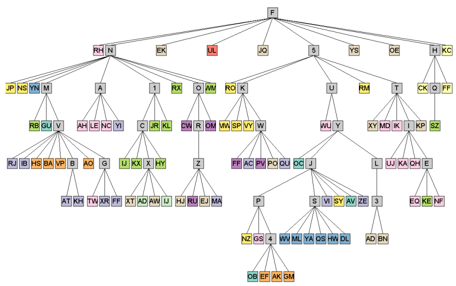
After responding to the three surveys, participants began the main study portion,



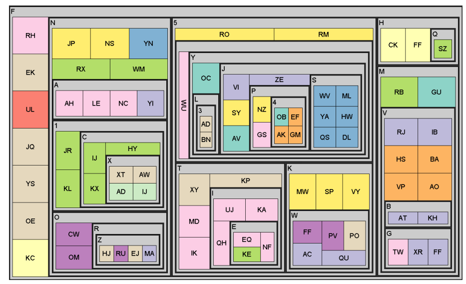
(a) Small node-link condition



(b) Small treemap condition



(c) Large node-link condition



(d) Large treemap condition

Figure 6: The visualizations used in the individual differences study.

which took place in a Java applet shown on the Mechanical Turk site. Participants first had a brief training procedure for both visualization types. In this phase, they were asked questions similar to those in the test phase, but were able to correct any mistaken responses until they got it right. The order in which they were trained on the two visualization methods was randomized. Once they answered all of the training questions in both visualizations correctly, they moved on to the test phase.

This phase consisted of four blocks. Each block consisted of a visualization, either treemap or node-link, depicting a separate dataset, and 16 questions that required the participant to consult the visualization for information about the data. These were yes-or-no questions, which the participants answered by pressing either q for "yes" or p for "no." The questions were of four types, and the sixteen questions in a block consisted of four groups of four questions of the same type. The question types

Table 3: Examples of the four types of task questions asked during the experiment, in either a *containers* or *levels* metaphor. The *containers* metaphor is thought to be compatible with a treemap visualization and the *levels* metaphor is compatible with the node-link visualization. Participants saw four versions of each of these questions, with different files or folders substituted, in each of the four study conditions, for a total of 64 task questions.

Containers	**Levels**
1. Does directory H contain a deeper hierarchy than directory P?	1. Does directory H have more levels under it than directory P?
2. Does directory W contain more subdirectories than directory H?	2. Are there more subdirectories under directory W than directory H?
3. Are there more files immediately inside directory R than directory F?	3. Are there more files immediately below directory R than directory F?
4. Are both file RV and file KH within directory R?	4. Do both file RV and file KH fall under directory R?

were worded in either a *containers* or a *levels* metaphor (Table 3).

Each question was first displayed against a blank screen. Once the user indicated that she had read the question by pressing a "Done" button, the visualization appeared and the user was given time to consult the visualization and answer the question by striking the appropriate key on the keyboard. We measured response time, reading time, and accuracy for each answer.

The visualization and the verbal metaphor of the questions varied within subjects, so that the four blocks were as follows: node-link and *levels* metaphor (NLL), treemap and *levels* metaphor (TML), node-link and *containers* metaphor (NLC), and treemap and *containers* metaphor (TMC). We consider the NLL and TMC blocks to have compatible visual and verbal metaphors, while the NLC and TML blocks have incompatible visual and verbal metaphors.

In order to measure the contribution of difficulty to the participants' performance on task questions, we varied the number of levels in the four hierarchy datasets. The

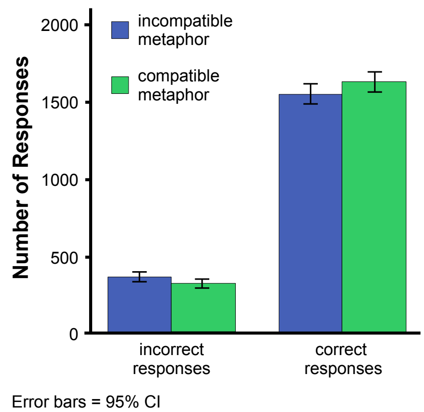


Figure 7: (a) There was an overall significant effect of question compatibility on response correctness. However, there was no such significant difference for participants with above-average spatial ability, participants who scored highly on the personality dimension of Openness, or participants who reported no verbal metaphor preference.

first two were simple four-level hierarchies, while the second two were larger eight-level hierarchies. We counterbalanced the four metaphor conditions across the blocks to correct for order effects and the potential interaction of difficulty and compatibility.

5.2.2 Results and Discussion

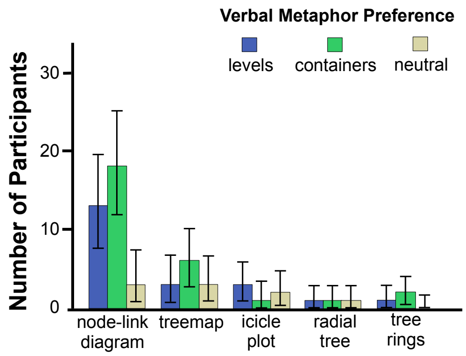
With 63 participants responding to 64 questions each, we received an initial total of 4032 responses. 155 cases with a response time of less than one second were assumed to be errors and were dropped from the final data, giving us a total of 3877 responses. This removal does not affect the significance of any of the tests reported in this section. Response times ranged as high as 101.4 seconds, but we did not consider any responses long enough to warrant dropping. The mean response time was 12.19 seconds ($S.D. = 9.53$) and the mean time to read a question was 3.5 seconds ($S.D. = 3.6$). The reading time is likely very low because participants would see four versions of the same question in a row, making it unnecessary to read anything but the specific files or folders being referenced.

We analyzed our results with a focus on how metaphor compatibility affected response time, reading time, and correctness of responses. We found significant effects from a number of the individual factors which shed light on how and why users experience metaphor interference.

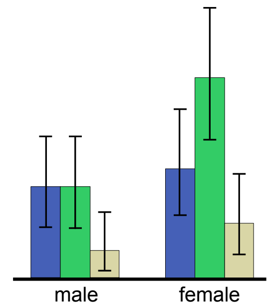
5.2.2.1 Compatibility and correctness

Using Pearson's Chi-Square, we found an overall significant effect of metaphor compatibility on correctness, $\chi^2(1,\ N = \ 3877) = 3.93$, $p < .05$, confirming our primary hypothesis that questions in a compatible metaphor are easier to answer (Figure 7). We did not find this effect in our previous study; however, we believe that the better control of question difficulty and condition separation in the current study may account for this difference. This effect is not influenced by the difficulty of the dataset, nor was there a significant difference between the treemap and node-link conditions.

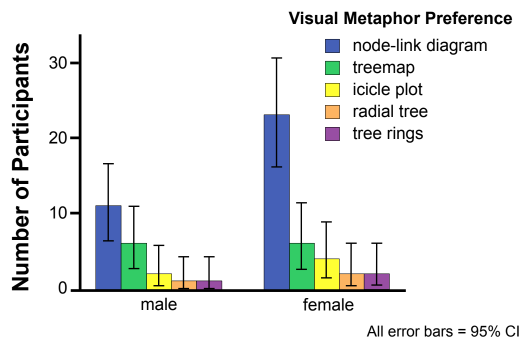
Several individual factors influenced the extent to which a participant showed this correctness effect. We divided participants into Low, Average, and High Spatial Ability groups based on their responses to the Form Board test. Low Spatial Ability participants were defined as those with a score lower than 86.5, or less than one standard deviation below the average as reported by the College Board [21]), and High Spatial Ability participants were those with a score greater than 163.1, or greater than one standard deviation above the average. The overall accuracy of the High Spatial Ability group (85%) was also significantly higher than those of the Average (79.8%) and Low (78.9%) groups, $\chi^2(2, N = 3877) = 20$, $p < 0.001$. However, unlike the



(a)



(b)



(c)

Figure 8: (a) While these differences are not significant, patterns of self-reported verbal metaphor preference can be seen among participants who ranked a given visual metaphor (Figure 5) as the best depiction of a hierarchy. (b) While women wore more likely than men to report a preference for verbal metaphors of *containers*, (c) they were also less likely to choose a treemap as the best visual metaphor for a hierarchy.

Low and Average groups, High Spatial Ability participants did not show a significant difference in correctness between compatible and incompatible questions.

Similarly, participants who scored highly on the personality dimension of Openness (defined as greater than 4.43, or one standard deviation above the population average [20]) showed no significant difference in correctness between compatible and incompatible questions. There was no significant correlation between spatial ability and Openness, suggesting that they are independent predictors of a user's ability to translate rapidly between conceptual metaphors.

5.2.2.2 Compatibility and Response Time

We did not find a significant effect of metaphor compatibility on response time, although a univariate ANOVA did find an effect of compatibility on reading time (the time it took a participant to read the task question), $F(1, 3795) = 12.99$, $p < .001$. This may indicate a priming effect, a common finding in psychology studies in which one stimulus (in this case, the visual metaphor) facilitates the processing of a subsequent related stimulus (a compatible verbal metaphor). This effect may account for some of the ambiguity of our previous study, in which we did not distinguish between reading and response time. As in the effect of compatibility on correctness, the difference in reading times between compatible and incompatible metaphors is not significant for participants in the High Spatial Ability group. Interestingly, participants in the Low Spatial Ability group tended to read incompatible questions faster, although this difference in speed was not significant. The participants in the Average Spatial Ability group most strongly showed the main effect of compatibility on reading time, $t(1484) = 3.519$, $p < .001$. The overall interaction between compatibility and spatial ability for reading time was significant, $F(2, 3791) = 5.24$, $p < .01$.

5.2.2.3 Metaphor Preference

We measured a participant's verbal metaphor preference using their ratings of how well statements described a sample hierarchy, as described in Section 5.2.1. To determine which verbal metaphor a participant preferred, we calculated their average rating for all statements worded in a given metaphor (either *levels* or *containers*).

If the average ratings for the two metaphor groups were equal, we considered the participant to have a neutral self-reported verbal metaphor preference; otherwise, she was said to have a self-reported preference for the higher-rated verbal metaphor. To determine visual metaphor preference, we simply took the highest-ranked depiction as the participant's self-reported preferred visual metaphor for hierarchies. We took these self-reported preferences as a measure of the participant's preconceptions of hierarchical structure prior to starting the study. While this is a simplified approximation, as it does not consider cases where a participant's preconceived structure is not one of the options or is some combination of visual metaphors, it can at least capture the user's relative comfort with the two visual metaphors used in the study and visualizations related to them.

We did find some non-significant patterns of association between self-reported verbal and visual metaphor preference (Figure 8(a)). While a Pearson's Chi-Square test of independence between verbal and visual metaphor preference is not significant, participants who ranked treemaps higher somewhat tended to prefer a *containers* verbal metaphor. Contrary to our hypothesis, we did not find any association between preference for the *levels* verbal metaphor and the node-link diagram as a visual metaphor. These findings may call into question our assumptions about which verbal metaphors are compatible with the visualizations we use, or suggest that the correspondence between verbal and visual metaphors is indirect.

There were some gender-related patterns in these measures of self-reported preference. Women were more likely than men to prefer *containers* metaphors (Figure 8(b)), although this effect is not significant. However, women also non-significantly tended

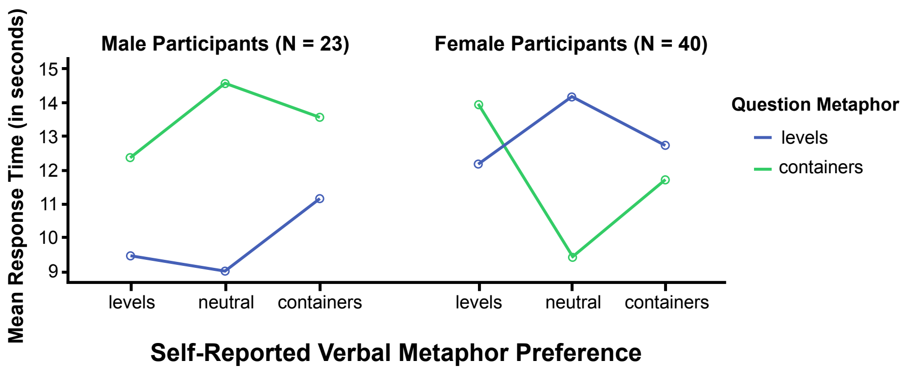


Figure 9: While we expected users to respond faster to questions in their self-reported preferred verbal metaphor, this pattern only emerged for women. Men responded significantly faster to *levels* questions no matter what their self-reported metaphor preference. Women who did not prefer one verbal metaphor over another tended to respond faster to *containers* questions.

to rank the treemap lower in their preferred visual metaphors (Figure 8(c)). We also found a significant effect that participants with higher spatial ability rated all verbal descriptions lower (that is, there is a negative correlation between spatial ability and overall verbal description rating, $R(57) = -0.29$, $p < 0.05$), suggesting a potential dichotomy between a comfort with spatial and verbal thinking.

5.2.2.4 Preference and Performance

The connection between self-reported metaphor preference and performance in the test portion was weaker than the effects of other individual differences, and showed a surprising gender effect. While we hypothesized that participants would generally perform faster for questions in their preferred metaphor, we found this effect only applied to women. An ANOVA on response time found a significant interaction between self-reported verbal metaphor preference, gender, and the metaphor of the question, $F(2, 3481) = 10.38$, $p < .001$. While women who reported a preference for one verbal metaphor over another had significantly faster response times in that metaphor, men had significantly faster response times on levels questions no matter what their self-reported verbal metaphor preference (Figure 9). However, we did find that participants with a self-reported neutral verbal metaphor preference did not show a significant compatibility effect on correctness, providing evidence for our hypothesis that a preconceived metaphor for hierarchies leads to a greater compatibility effect.

We did not find a significant effect of self-reported visual metaphor preference on correctness or response time across the two visualization types. That is, users who ranked treemaps the highest out of all visual metaphors did not respond more correctly or faster to questions in the treemap condition, and likewise for users who ranked node-link diagrams the highest. Similarly, although women ranked treemaps lower consistently, there were no significant gender differences in correctness or response time in either the treemap condition or the nodelink condition.

5.2.2.5 Other Factors

There was no significant difference in accuracy between the node-link and the treemap conditions, although response times in the treemap condition ($M = 11.6$, $S.D. = 8.5$) were significantly faster than in the node-link condition ($M = 12.7$, $S.D. = 10.4$) by a small amount, $t(3875) = 3.496$, $p < .001$. There was no significant difference in accuracy or response time between the two verbal metaphors. Unsurprisingly, responses to the small four-level hierarchies were significantly faster ($t(3875) = 36.8$, $p < .001$) and more accurate ($\chi^2(1, N = \ 3877)$ 4.4, $p < 0.05$) than responses to the eight-level hierarchies. We did not find any significant differences in the compatibility effects between the small and large hierarchy conditions.

These results generally confirm and clarify our initial findings that conflicting metaphors can cause decreased performance on simple visualization tasks. The fact that reading time showed a significant effect while response time did not suggests that the conflict may lead to slow processing of the problem, rather than slow processing of the visualization. This may indicate that the visualization is a stronger structural cue than the wording of the task question, or that structural has a priming effect that slows down the processing of conflicting stimuli but not the actual data reading process. Finally, we found that flexible thinking, defined various ways, is a good predictor for the ability to switch between structures. This clarifies the individual differences in the previous study and shows that structural translation is a major part of the visualization process.

5.3 Implications of Structural Conflicts

The fact that conflicting visual and verbal metaphors can affect performance at simple information extraction tasks in a visualization is strong initial evidence that visual information structure is a significant part of visualization use and that metaphors make a good framework for understanding the effects of visual structure. At the same time, the results of these studies begin to add nuance to our original model of the role of structure in infovis.

An obvious point on which our model must be clarified is that different people react to visual structure differently. These findings show that some process of structural translation must be taking place in some cases, but that for some people this is easier than others. It is also possible, given our results, that those participants who do not show an effect of conflict on performance are not experiencing structural translation at all, but are understanding information from a variety of structures in a more abstract form.

Either way, the patterns suggest that those who do not experience structural conflicts are the more flexible thinkers. This is partially reminiscent of findings in verbal metaphors showing that expert users rely less on metaphors in problem-solving; perhaps the flexible thinkers required less structure to understand the problems they were given. However, this does not explain the finding in the first study that those with higher conflicts also had higher accuracy, suggesting that thinking through the difficult structural translation had a positive effect on some users' understanding of the tasks.

In general, however, these findings suggest that complex thinking involves switching between structures, and that easier structural switching shows a facility with such complex thinking. Users with high spatial ability, for instance, may find it easier to translate between the implicit spatial structure of the task question and the spatial structure of the visualization, while those with more imaginative thinking styles (indicated by the Openness personality dimension) are more comfortable thinking in several modes at once.

An intriguing possibility suggested by the finding that structural conflicts are associated with accuracy is that conflicting structures may encourage deeper consideration of the visualized information. That is, those participants who took the most time translating between structures were less efficient, but ended up learning the tasks better than those who attempted to answer questions without reference to information structures. This may not have been a factor in the second study, which had less variance in task difficulty. This suggests that structural conflicts in a visualization may not necessarily be a bad thing, possibly sacrificing efficiency for the sake of deeper understanding.

The gender effects in self-reported metaphor preference were intriguing, and may be related to known gender differences in spatial and verbal abilities. The implication that men and women may approach visualization with different preconceptions and thinking styles certainly bears further study. One potential factor that we did not consider in this study is verbal ability. It is possible that verbal comprehension skills may play a part in the influence of verbal metaphors on a user's thinking process, and it is a common finding that women have higher scores than men on tests of verbal ability [31].

While these findings establish the importance of visual information structure and argue for an underlying similarity between visual and verbal metaphor, they are limited in what they tell us about how visual structure is interpreted. The study designs assume that treemaps embody a *containment* metaphor and node-link diagrams embody a *levels* metaphor, but in reality we have no way of knowing for certain that this is how people interpreted these structures. For example, other researchers commenting on this work have noted that they see treemaps as objects stacked on top of each other, seen from above. This would obviously confound the expectation that questions in a *containers* metaphor would be more compatible with the treemap.

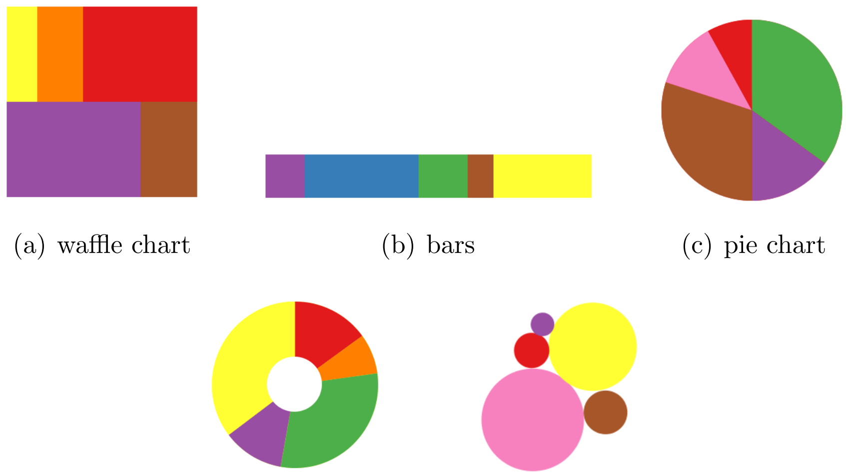
So far in this work, I have approached visual structure from a high level. While this approach is appropriate as a first attempt at determining the viability of the visual metaphor model, it does not lend itself to understanding the details of how people perceive and use visual structure in a visualization. In order to do so, it is necessary to look at visual structure from a lower level: what are the basic elements of visual structure, and how do people metaphorically interpret them?

CHAPTER 6: HOW VISUAL STRUCTURE IS INTERPRETED

The next stage in this research is to understand more concretely how people turn visual forms into information structure. In the initial studies on visual metaphor effects on visualization, I explored metaphors for hierarchies that function at the *structural* level (as explained in Section 3.1. That is, the metaphors *a hierarchy is a series of levels* and *a hierarchy is a set of nested containers* are both extensive mappings between complex domains. To understand visual structure at a more basic level, it is necessary to move to less elaborate orientational and ontological metaphors. This means using less complex visualizations and studying how people turn simple shapes and visual patterns into information structure.

6.1 Metaphorically Interpreting Visual Design

In order to study the effects of visual structure at a lower level, it is necessary to begin making predictions about what elements of a visualization can carry structural information. Predicting the effects of visual structure would make it possible to explain and control biases in visualization reading such as those we found in Chapter 5 as well as the more serious presentation effects on decision-making found by Elting et al. [22]. By isolating these elements and how they affect perceptions of information, we can begin to develop theories to predict what aspects of a visualization will carry structural information and what the nature of that information will be.



(a) waffle chart (b) bars (c) pie chart

(d) donut (e) bubble chart

Figure 10: The five chart types used in the study.

We hypothesize that design elements that carry implicit physical information, such as borders, connectedness, and background shapes, can have significant and consistent effects on how a user interprets a visualization. Specifically, we believe that elements of visual design can be shown to affect subjective responses to entities represented in visualized data in a reliable and systematic fashion. We further hypothesize that while these effects will be present and to some extent consistent across visualization types, they can also be generated by the visualization type itself, and design elements will affect different visualizations in different ways.

6.1.1 Procedure

In order to test our hypotheses, we designed a study [66] to test the effects of visualization design elements on semantic judgments of data in a simple context. This study was meant to test whether a particular set of design elements indeed have a significant effect on semantic judgments, and to what extent these design elements affect various types of simple information visualizations.

We recruited 43 participants via Mechanical Turk. Participants performed the study online and were paid a base rate of \$0.20 for their work, which took at most twenty minutes. They were told that giving especially helpful responses (defined as detailed, thoughtful, and clearly written) would yield a bonus payment of \$1.00. Twelve participants ultimately received this bonus.

One participant was dropped from the study and not compensated due to a clear attempt to cheat the system by entering the same response for every question, leaving a total of 42 participants. (Participants were warned in the study instructions that attempts to submit the work assignment without data would be rejected.) Of the remaining participants, 25 were female (59.5%) and 17 were male (40.5%). Participant age ranged from 21 to 62, with an average age of 36.4.

Participants viewed a series of twenty charts which were described as representing fictional companies. Each company was divided into six departments, with the pieces in a chart representing departments sized according to the department's relative expenditures for a fiscal year. These proportions were the same for all twenty charts, but the order and coloring of the departments in each chart was randomized to conceal this fact. Although we cannot know for certain whether this was successful, only one participant's comments indicated that he realized that the proportions were all the same. The colors were derived from a categorical color scale from ColorBrewer [12]. This "company/department" description was used because it is a largely abstract concept that can lend itself to different conceptualizations, but is still familiar enough that participants could easily interpret data about it.

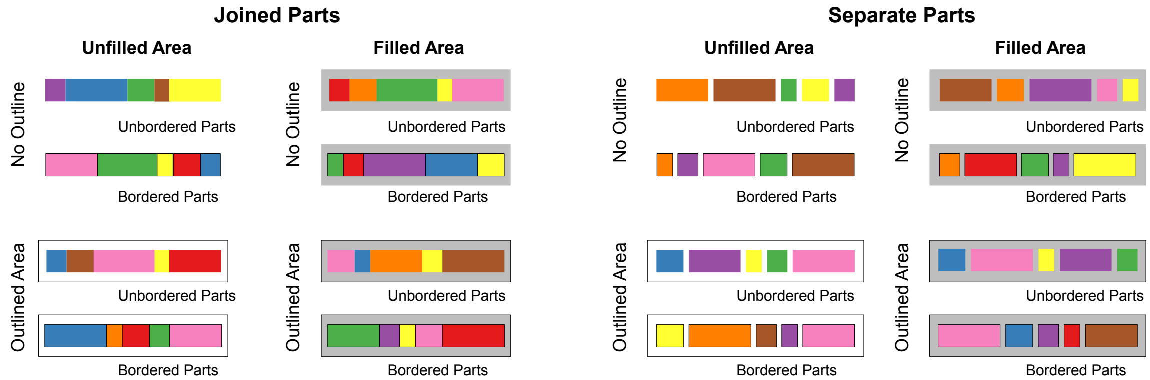


Figure 11: The stacked bar chart in each of the sixteen design configurations used in the study. These sixteen configurations represent every combination of the four binary design element variables we chose.

We used five types of charts to display the relationships of the departments to the company: a pie chart, a waffle chart (or one-level treemap), a horizontal stacked bar chart, a donut chart, and a bubble chart (Figure 10). These visualizations were chosen to represent a range of shapes and relative familiarity, while all being simple enough to evaluate quickly.

Having chosen this set of visualizations, we systematically altered them according to four design dimensions. These dimensions represented the presence or absence of some element we hypothesized to have semantic value. The elements were:

Filled area. In charts with a filled area, a gray background was visible behind the main chart which mimicked its overall shape but at a larger size. For the bubble chart, which is not space-filling, we used a circle as the filled area.

Outlined area. In charts with an outlined area, a black contour was drawn around the chart in the same shape as the main chart but at a larger size. For the bubble chart, we used a circular outline.

Bordered parts. In charts with bordered parts, black contours were drawn around each of the individual pieces.

Joined parts. In charts with joined parts, pieces were connected to one another in the manner most natural for a given visualization type. In separated parts conditions, the view was "exploded" so that a small amount of space was visible between parts.

With two possible states for each of these four variables, there were a total of sixteen design configurations for each of the chart types. Figure 11 shows these sixteen configurations applied to the bar chart.

These design elements were chosen because outlines, color areas, and connectivity are among the cues that visualization researchers such as Ware [59] and Tversky et al. [56] consider to be meaningful visual primitives. Additionally, we hypothesized each to have a unique semantic effect on the dynamics of the visual representation. We expected a filled area to suggest a stable foundation, the outlined area to suggest a limit or fence around the entire company, bordered parts to suggest limits on individual pieces, and the joining of parts to suggest connections between pieces.

To test these hypotheses, we developed a list of ten semantic variables which could describe a simple dataset of the kind we presented in structural terms. On a scale of one to five, we asked participants to rate how much the company presented in a chart was likely to be *stable, complete, controlled, inflexible, rigid, structured, isolated, unified, well-organized*, and *good place to work*.

While these variables were primarily chosen to represent a broad range of abstract structural qualities a company might have, they can also be interpreted as representing a range of dynamic physical properties. For example, *rigid* and *inflexible* may imply that a scene cannot be simulated as changing very much. *Stable* and *controlled* suggest that a figure is well supported and will not move without additional forces. *Unified* and *complete* may suggest that parts are seen as attached to one another and will move as a unit, while *isolated* may suggest that parts will move independently. *Structured* may imply that the scene represents a connected unit with dynamics in balance. The last two semantic variables, *good place to work* and *well-organized*, are more subjective and may represent combinations of the other variables.

We varied two of the four design elements between subjects (filled areas and bordered parts) and two of the elements within subjects (outlined areas and joined parts). The purpose of varying some elements between subjects was to reduce the number of charts a participant saw in order to avoid fatigue. We hypothesized that the design elements we chose as between-subjects variables were less likely to have an effect than the within-subject variables, although this did not entirely prove to be the case.

In the first part of the study, participants saw a series of twenty charts described as representing fictional companies. Participants were told that segments in a chart represented the departments of the company, and that the size of each segment represented the amount of spending by that department over a fiscal year. The ten semantic variables were presented below each chart, in an order which was randomized for each participant but kept consistent throughout a single participants' progress. Participants were told to rate each company on these semantic variables to the best of their ability based on the information in the chart. The charts each participant

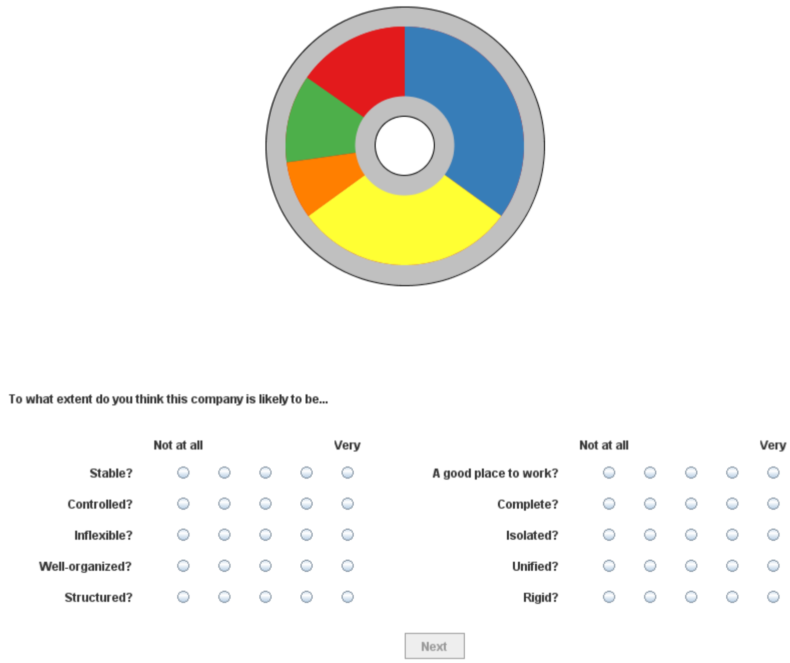


Figure 12: The study interface as seen by participants.

saw included four versions of each of the five chart types, varied on the two within-subjects variables of *outlined area* and *joined parts*. These charts were presented in random order. After the participant rated each chart on all ten variables, she clicked a button to continue to the next chart.

In addition to this main part of the study, we wished to give the participants an opportunity to explain their ratings in more detail so we could better understand how they viewed the charts semantically. We also wanted to make sure that participants felt comfortable rating the companies based on the charts and were able to give reasons for their ratings. After the first part of the study, we chose four pairs of charts to present to the participant. These pairs included two pairs in which the charts were visually similar (that is, were mostly the same on the visual dimensions) but were rated very differently by the participant, and two pairs in which the charts

Table 4: Factor loadings for the two factors we derived from the original ten semantic scale items, using a principal component analysis with Varimax rotation.

Scale Item	**Factor 1 (Balance)**	**Factor 2 (Strictness)**
Complete	.795	.053
Controlled	.530	.559
Good Place to Work	.633	-.368
Inflexible	.141	.852
Isolated	-.250	.530
Rigid	.211	.852
Stable	.819	.057
Structured	.752	.331
Unified	.753	.073
Well-organized	.823	.062

were visually different but rated similarly. We calculated the difference in ratings for every pair of charts by taking an average of the absolute difference between their ratings on each of the ten semantic dimensions. During the second phase of the study, we presented the two charts in each pair to the participant side by side and provided a text box in which she was asked to explain why she rated the charts either similarly or differently.

Once the participants finished these two phases of the study, they were asked to provide their gender and age in an additional form on the Mechanical Turk site.

6.1.2 Results

Since our ten items in the scale we employed likely show a good deal of semantic overlap, the first step in analyzing these results was to reduce the number of dimensions using factor analysis. We employed a principle components analysis with varimax rotation and selected the resulting factors with an eigenvalue greater than one, which produced two factors. On examination of the factor loading (Table 4), the first factor seems to encompass the scale items that are more positive and suggest stability and good organization, while the second factor encompasses the items that are more negative and suggest rigidity and oppressiveness. However, since the written responses by participants show some variation between subjects as to which of the scale items were considered negatively or positively, I will refer to these factors more neutrally as “Balance” and “Strictness” in the subsequent analysis.

The effect of our overall model (a 2x2x2x2x5 design, with the four design elements and chart type) on these factors was first assessed with a multivariate analysis of variance (MANOVA) using Wilks’ Lambda as the test statistic . This analysis found a significant main effect of chart type, $F(8,31) = 3.04$, $p < .05$, $\eta^2 = .44$. There was also a significant effect of joined parts, $F(2,37) = 7.11$, $p < .01$, $\eta^2 = .28$. In addition, there were significant interactions of filled area by bordered parts ($F(2,37) = 3.82$, $p < .05$, $\eta^2 = .17$), outlined area by joined parts ($F(2,37) = 3.94$, $p < .05$, $\eta^2 = .18$), and chart type by outlined area by filled area ($F(8,31) = 2.51$, $p < .05$, $\eta^2 = .36$). There was also a marginally significant interaction between chart type and joined parts, $F(8,31) = 2.21$, $p = .054$, $\eta^2 = .36$. Results that were significant in the multivariate analysis were further analyzed with univariate analyses of variance (ANOVA) for the two factors.

6.1.2.1 Main Effects

A summary of the main effects of the four design elements on the factors of Balance and Strictness is presented in Table 5. The only design element that produced a significant main effect in the MANOVA was joined parts, which a univariate repeated measures ANOVA found to be significant on the factor of Balance, $F(1,39) = 12.18$,

Table 5: Means in the two derived semantic factors for each of the four design elements we studied. Only the variable of Joined Parts had a significant main effect on the factor of Balance.

Factor	Group	Balance		Strictness	
		M	S.D.	M	S.D.
Filled Area	yes	-.032	.12	.004	.11
	no	-.067	.12	.106	.12
Bordered Parts	yes	-.071	.12	.125	.12
	no	.043	.11	-.086	.11
Joined Parts	yes	.160	.09	.073	.09
	no	-.188	.10	-.035	.09
Outlined Area	yes	-.030	.09	.022	.09
	no	.002	.08	.016	.09

Table 6: Means in the two derived semantic factors for each of the five chart types.

Chart Type	Balance		Strictness	
	M	S.D.	M	S.D.
waffle	.179	.11	.301	.11
bars	.085	.13	.204	.13
pie	.116	.08	-.045	.10
donut	-.002	.09	-.093	.12
bubble	-.447	.14	-.271	.12

$p < .001$, $\eta^2 = .24$. That is, charts with joined parts ($M = .16$, $S.D. = .09$) were rated significantly higher on this factor than those with separated parts ($M = -.19$, $S.D. = .09$).

In addition, we found a significant main effect of chart type on both Balance, $F(2.73, 39) = 8.32$, $p < .001$, $\eta^2 = .18$, and Strictness, $F(2.43, 39) = 6.38$, $p < .001$, $\eta^2 = .14$. (Since the variable of chart type did not meet the assumption of sphericity for either factor, we employed a Greenhouse-Geisser correction on these ANOVAs.) The means for each chart type on these two factors are summarized in Table 6. In general, these results show that the bubble chart is rated as much less Balanced than the other charts, meaning it was seen as less stable, unified, complete, and well-

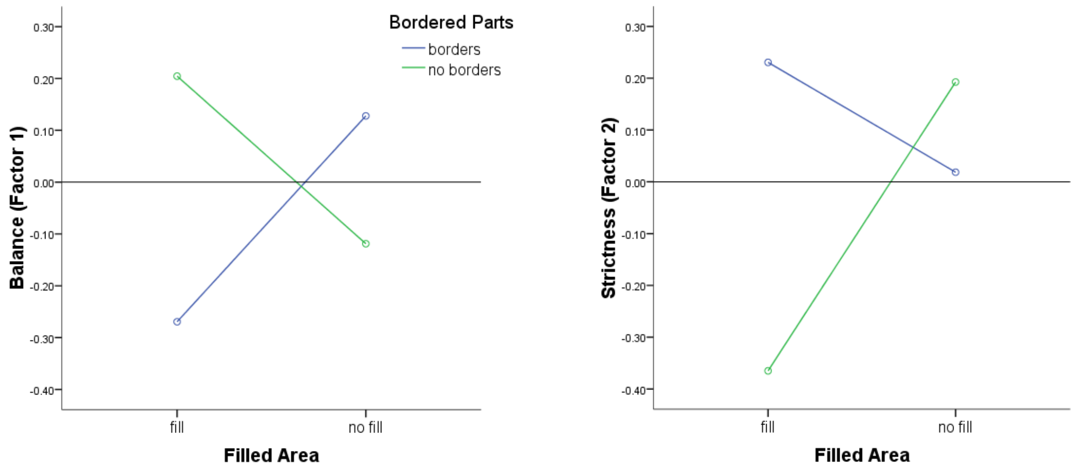


Figure 13: We found a significant interaction between the design factors of filled area and bordered parts. Balance increased when only one of the factors was present, but decreased when both or neither were present. Strictness was increased when both or neither were present, but was much lower when there was a fill but no part borders.

organized. Bubble charts were also rated as less Strict than the other chart types, and the rectangular charts–i.e., bars and waffles–were rated as more Strict than the predominantly circular charts, meaning they were rated as more rigid, inflexible, structured, and controlled.

6.1.2.2 Interactions

In addition to these main effects, we also found several significant interactions in our overall model. Repeated measures ANOVAs found the interaction between filled area and bordered parts to be significant for both Balance ($F(1, 39) = 4.95$, $p < .05$, $\eta^2 = .12$) and Strictness ($F(1, 39) = 5.28$, $p < .05$, $\eta^2 = .12$). These interactions are illustrated in Figure 13. Either a filled area or bordered parts on their own seem to increase perceptions of Balance, but the presence of both elements or neither element creates the perception of less Balance. This suggests that the positive effects these elements have on perceptions of organization interfere with one another in some way.

Table 7: Means for each of the five chart types across the design variable of joined parts.

Chart Type	Joined Parts	Balance		Strictness	
		M	S.D.	M	S.D.
waffle	joined	.406	.13	.370	.11
	separated	-.049	.13	.231	.14
bars	joined	.140	.15	.202	.14
	separated	.029	.14	.206	.14
pie	joined	.340	.12	.121	.13
	separated	-.109	.12	-.211	.11
donut	joined	.295	.11	-.132	.12
	separated	-.299	.13	-.054	.14
bubbles	joined	-.382	.14	-.197	.13
	separated	-.512	.16	-.346	.13

The effect of the interaction on Strictness adds some nuance to this interpretation. A chart with both a filled area and bordered parts is rated as highly Strict, as is a chart with neither element, which is complimentary to the Balance effect. However, while a chart with borders and no filled area is rated neutrally on the factor of Strictness, a chart with a filled area and no borders is rated as much less Strict. This suggests that a filled area produces a perception of flexibility which is somehow tempered by the presence of borders around parts.

The interaction of outlined area by joined parts was significant only for the factor of Balance, $F(1,39) = 8.08$, $p < .01$, $\eta^2 = .175$. This interaction seems to arise from the fact that the difference in Balance between a chart with joined and separated parts is larger when there is no outline around the chart area. This offers the intriguing possibility that the perceived instability of the “exploded” charts is mitigated when there is a boundary limiting the perceived motion of the pieces.

In addition to interactions between design elements, we also found minor but significant interactions between chart type and joined parts for both the factor of Balance ($F(4, 39) = 3.56$, $p < .01$, $\eta^2 = .09$) and Strictness ($F(4, 39) = 2.51$, $p < .05$, $\eta^2 = .06$). The means for these conditions are summarized in Table 7. Generally, the perceived Balance of a bar chart or bubble chart is less affected by whether pieces are joined or not than that of other chart types. Additionally, pie charts show a greater loss of Strictness when parts are separated than do the other chart types. As with the factor of Balance, bars are largely unaffected by separation for the Strictness factor, and donut charts show a slight trend in the opposite direction, with separated charts perceived as more Strict on average than joined ones.

Finally, there is a three-way interaction of chart type, outlined area, and filled area for the factor of Balance, $F(4, 39) = 2.98$, $p < .05$, $\eta^2 = .07$. This difference seems to be mostly attributable to the fact that for waffle charts alone, an outline increases the perception of Balance only when there is no fill, and vice versa. This recalls to some extent the interaction of filled area and bordered parts, and suggests that outlines and fills, combined with an already quite rigid visual structure, can lead to a more negative impression.

6.1.3 General Discussion

It is clear from the quantitative results of this study that design elements in a simple visualization context can have significant and to some extent consistent effects on a user's semantic evaluation of data. In addition to the simple semantic ratings, we also attempted to analyze the reasons for these ratings in the second part of our study, in which participants were asked to explain selected ratings. These explanations can shed some light on what design elements mean to users. There were no significant patterns to the types or configurations of charts automatically chosen for comparison, so participant comments covered a wide range of conditions. In general, these comments tended to reinforce our quantitative results and provided anecdotal evidence that a tendency to interpret charts as dynamic physical scenes leads to the patterns of semantic ratings that we found. They also demonstrate that our users were easily able to explain their ratings in almost all cases, suggesting that the task, while unusual, was understandable; only one of our 42 participants reported having difficulty rating the companies based on the charts. While many comments were minimal or did not address structural properties (for example, comments in which a participant preferred one chart to another because the colors were more pleasing), those which were more elaborate did tend to talk about dynamic and physical qualities of the chart.

6.1.3.1 Physical Dynamics of Design

One common theme throughout these explanations was a treatment of both design elements and chart types as offering various potentials for movement or communication. For example, two separate participants explained that they considered the donut chart less stable because it seemed like it might "roll away." This kind of analysis also seemed to underly the evaluations of the bubble chart as unstable and uncontrolled, with participants describing this chart as "floating bubbles that were barely contained within the area" and "scattered." (It should be noted that apart from a mention of the pie chart in the initial instructions, we did not name or label the charts in the study, so the participant who described the bubble chart as "bubbles" did so spontaneously.) These comments suggest that at least some participants were implicitly applying gravity to the charts, and found the bubbles disconcerting because they so strongly seemed to violate the constraints of gravity.

This kind of description may also help to explain why perceived Strictness was seemingly reduced with a filled area. One participant describing a bubble chart on a filled area said that it looked "as if the parts were placed randomly, with room to move them around," and another that it seemed "as if the company doesn't quite know how big it is." Of the five visualization types we presented, the bubble chart was the one with the least constraints on placement of parts; in the other four cases, the pieces are filling a predefined space of one shape or another. The arbitrary nature of the bubbles' placement, then, may be highlighted when there is a clearly defined space in which they can "move." In contrast, one participant described departments in the space-filling waffle chart as having "no room to move around."

A similar sense of physical potential seemed to inform participants' ratings of the different design elements. For example, a common observation was that charts without joined pieces seem to be "flying apart" or "exploding." Since the pieces in this case are not supporting each other, the scene may be interpreted as being in a state of motion. It is therefore more unstable, since the pieces have not yet come to rest. Another possibility is that participants see a joined chart as a natural state, so that separation between pieces implies movement.

6.1.3.2 Design Element Interactions

Aside from the expected physical simulations pointed to by these comments, another common theme was that different design configurations allowed for different amounts of communication between parts. This theme sheds light on the interaction between filled area and bordered parts. Describing a pie chart and a bubble chart that she rated similarly despite their differing on all dimensions except for filled area and bordered parts, one participant wrote that "they both represent difficulties in communication within the organization." Since the filled areas and bordered parts condition places two types of barrier between or around pieces, it may be seen as allowing less communication between parts and therefore a worse place to work.

A similar kind of analysis based on combined design features is hinted at by a participant who compared two charts with part borders, one with joined parts and one without. This participant described the company with separate parts as being "closed" and the one with joined parts as "not so closed," suggesting that the boundaries around parts are seen as less rigid when parts share borders. This participant went on to describe the company with joined parts as having "a more controlled flexibility."

Finally, the combined effect of filled areas and joined parts was referred to by two participants, although they seemed to offer two entirely different reasons for the same assessment. One, comparing two similarly rated charts with filled areas and separate parts, wrote, "Both of these charts have so much gray area between the departments." This suggests that filled areas and separate parts can be viewed negatively due to the presence of a visible barrier between parts. However, another participant who was comparing a chart with a filled area and separate parts (on the left) to one with a filled area and joined parts (on the right) wrote, “The company on the left is off its foundation (the gray circle), whereas the company on the right is centered on its foundation.” (Both charts were, in fact, centered within their filled areas.) This offers the alternate possibility that the movement suggested by separated parts may create a sense of precariousness if the user views the filled area as a foundation on which pieces rest. Interestingly, these two sets of comments imply that a chart may lend itself to different interpretations depending on whether the user perceives gravity as moving downwards or moving into the screen; that is, whether she sees herself as looking at a side view or a top-down view of the scene.

The physical interpretations of chart elements provided by participants suggest the potential usefulness of these kinds of dynamics as a framework on which to hang a theory of how design elements contribute to visual structure. However, the apparent differences in how participants interpreted these physical properties suggested in the last example make clear that reliably defining these mappings between visual elements and physical properties is by no means a trivial process.

6.1.3.3 The Extent of Elaborations

While users’ comments suggested a strong influence of physical simulation on their semantic perceptions of the charts they saw, it could be argued that this influence does not go beyond simple visual organization to affect how participants actually think about the data. Most comments, such as those already quoted, tend to focus on properties that could apply equally well to either visual structure or the more abstract structure of a company; for example, balance, barriers, and movement are all common themes. This relatively direct metaphorical application of visual properties to abstract properties is largely what we expected to find in the user comments.

However, participants' comments often went beyond this simple metaphorical mapping to make elaborate inferences about the company's behavior and management style in terms that did not obviously map to visual properties of the chart. This was especially common in descriptions of companies visualized with a bubble chart, which various participants found "more organic and cooperative," "fun and open," "easy to get along with," and "open source." At the same time, one participant stated that they could not treat the company portrayed in a bubble chart seriously. The waffle chart also elicited a number of emotional responses in the opposite direction, and was described as "bulkier," "organized," and "regimented." These emotional responses to the waffle and bubble charts recall the fact that these were also the types that tended to receive the most extreme ratings on the semantic variables (Table 6). Since these are probably the two least familiar charts presented in our study, these results suggest that extensive elaboration is a greater factor in novel charts, whereas highly conventional representations such as pie charts may require less active interpretation on the user's part.

There were several cases in which these elaborate inferences went so far as to suggest that participants were able to imagine entire stories about the companies based on the simple charts they viewed. In describing a waffle chart, one participant wrote, "Going by the rules is the most important thing in this company and to violate them can get you in serious trouble." Another, describing the similarities between a donut and bubble chart, wrote, "They have some rules but they are mainly focused on encouraging people to do their best in terms of reaching a mutual goal. They don't want to stifle creativity, they want to encourage it." Eight of the 42 participants wrote at least one description that involved this kind of storytelling about the depicted companies. While these usually were based on the chart type, one participant elaborated upon the company based on the presence of a outlined area: "The graph on the left is encompassed by an extra circle—I took this to mean that there was some kind of higher up that kept all the smaller parts in line." That participants made these kind of imaginative elaborations at all, and that there seem to be some commonalities among them, is striking in itself, and suggests that visual structure can in some cases lead not only to minor semantic responses but to full-fledged inferences about the data.

6.1.3.4 Color Dynamics

A final trend in user comments suggested another factor that we did not directly consider in our initial study design. The colors of individual pieces were randomly generated at the point each chart was displayed, and we did not record these color combinations. However, several comments suggested that the arrangement of color affected how some users perceived the weight and balance of the charts. One participant said that a red segment in one pie chart looked larger than a blue segment in a second pie chart, even though she realized they were the same size. Another participant wrote that "the two largest sections, in brown and muddy green, recede a bit,

Figure 14: The colors in our study ranked by luminance ($L*$ in CIELAB color space) from darkest to lightest. Perceptually darker colors are seen as heavier by viewers, an effect which was reflected in our participants' comments.

making them seem more balanced by the red, vibrant blue and pink—that is, the two bigger sections have colors that make them seem less dominant." These comments may reflect the established finding that darker colors are perceived as being heavier than bright ones [46], and the red color used in our study was in fact the color with the lowest luminance value.

The different weights perceived in pieces with different colors may have affected how participants viewed the dynamics of the charts. For example, the comments about color balance may reflect the perceived center of gravity of an object. Also, while describing a donut chart with separated parts, a participant wrote that "the pieces appear to be flying apart, especially the dominating bright red of the largest slice." This implies that the heavier color is perceived as giving extra velocity to the implied movement of the piece. While these comments point towards a role for color in interpreting the dynamics of a visual representation, the fact that we did not record this information unfortunately makes it impossible to interpret this trend in the current study.

The broad trend in these results and the feedback of participants is that elements of visual design carry structural information when they suggest certain physical dynamics or constraints on information. This finding points the way forward to a more grounded theory of how visual structure is interpreted by a viewer.

6.2 Implications of Design and Visual Structure

We have provided theoretical background and experimental evidence that the structural elements of a visualization design influence how users interpret the meaning of information. We further argue that these dynamic interpretations guide both inference and a user's perception of how a visual representation can be manipulated. This theory has a number of significant implications for the design of interfaces that use information visualization.

The first implication is that a better understanding of the semantics of design elements would make it possible to exploit their effects as is appropriate to the task. Rather than treating design as decoration or distraction, it could be used consciously to suggest global attributes of the dataset or to communicate interpretations of data in a collaborative context. While this may not be a novel idea to the design community at large, it is less intuitive in information visualization, where design elements such as those used in our study have traditionally been considered irrelevant at best. The fact that the presence or absence of these elements can actively influence interpretations of data, occasionally to a high degree of elaboration, suggest that design cannot be simply ignored or minimized in information visualization contexts. Every design choice is a choice about how the data will be interpreted.

A common theme in our results and user comments is that different design configurations suggested different levels of movement and freedom. The waffle charts and bar charts were seen as rigid and fixed; charts with filled background areas suggested more flexibility. These findings may have practical implications for designing interactive visualizations and other types of interfaces. Charts and visualizations with designs that imply flexibility and movement may suggest more potential points of interaction to the user. Rectangular, space-filling designs (such as treemaps) may be better suited to cases where limited interaction is needed, while a more open design with a clearly delineated space in which pieces can move may encourage users to interact more extensively.

The idea of implied dynamics may in general provide a framework for building intuitive interactions into novel user interfaces. How to make possible interactions in a novel system discoverable is a general problem in human-computer interaction, and is sometimes addressed with the idea of perceived affordances [47]. That an abstract scene such as a software interface may inherently present a set of implicit physical properties and forces suggests that designers could visually play up certain dynamics in order to guide a user's understanding of what can be done. For example, items that can be moved can be made to appear lighter, while items that cannot be changed can be made to look heavy and rigid.

Implied dynamics of visualizations may also constrain design in certain ways. Many of the negative comments made about the bubble chart and the various exploded charts (that is, those without joined parts) suggest that the lack of support and apparent violation of perceived gravity in these visualizations makes them seem chaotic and disorganized. This may or may not be a problem for a given task, but in the information visualization context it may have the undesirable effect of distracting from whatever organization or structure the data actually has. If implied dynamics are indeed an important part of visualization perception, then any visual representation that strongly implies motion, such as an exploded chart, should be used with care.

Another practical use for our findings is that they suggest certain visualization types which are more or less susceptible to the effects of design. In general, we found that the design elements we used had effects that were consistent across a variety of chart types, suggesting that they have basic effects on perception independent of their context. However, we also found a non-significant trend for ratings of the bubble chart to be more design-sensitive and those of the pie chart to be more robust to design changes. This may be an effect of the relative familiarity of the two chart types, or it may speak to a more essential difference between their structures. Studying the design robustness of a given visualization would provide valuable information in the evaluation process and help in making decisions about visualization use.

This study provides initial evidence for a model of visual structure interpretation based on implied dynamics. By supporting this model, we can begin to add detail and nuance to our overall model of information structure in visualization. If this model is a reliable one, it could be used to predict how a given visualization will be interpreted, where meaningful differences between visualization methods may lie, and why a given visualization method is successful or unsuccessful for a particular task. This is an important step in moving beyond the simple visual encoding model. However, while evocative, the comments of our participants are not yet enough evidence on which to hang our entire model. In order to show that implied dynamics are a good model for how visual forms become visual metaphors, it is necessary to demonstrate not only that people apply dynamics to marks in a visualization, but also that the dynamics they perceive are related to the conceptual relationships they ultimately derive.

CHAPTER 7: VISUAL STRUCTURE AND IMPLIED DYNAMICS

In Chapter 5 I demonstrated that switching between conflicting metaphorical structures can complicate visualization use, and in Chapter 6 I showed that people can derive meaning from structure and apply it to data when prompted. These findings show that structure can be read by users of a visualization and that this reading of structure takes place during information extraction tasks. However, they do not prove by themselves that the structure people derive from a visualization influences the way they interpret data. At this point, an argument can still be made that all these structural effects are happening parallel to the core process of visualization use, which is decoding data values at the level of visual marks.

I argue, however, that this perspective is fairly naïve with respect to what we know about human cognition. The fact that simple changes in context can drastically change how people respond to the same information is very nearly a general principle in cognitive science, most famously demonstrated by Tversky and Kahneman's finding [55] that changes in the framing of a situation, such as wording a risk assessment in terms of lives lost versus lives saved, can predictably change participants' choices about the best outcome, even when the situation described is exactly the same. Assuming that the information in a visualization can be read and used entirely independently of the context presented by the visual structure seems more reckless than allowing that structure may have an effect on low-level data reading.

In order to test this hypothesis, however, we need a model to predict what effects structure might have on individual data points. The results of the previous study suggest a direction for such a model in the form of implied dynamics.

7.1 Implied Dynamics in Visual Metaphors

While my research up to this point has helped to establish that visual structure is important, it has not gone into the level of detail needed to predict how people read visual structure. This is like knowing *that* color scales affect perception, but not knowing *how*; it gives us enough information to be concerned but not enough to do something about it.

The first step is to understand what the atomic elements of visual structure are and why. That is, what visual forms do people see as carrying information about structural properties of data, such as how pieces fit together as a whole, how parts of the data relate to each other, and what patterns of behavior and inferences can be made about them. The results presented in Chapter 6 suggest some possibilities for these visual forms, including borders, outlines, and connection; these are, of course, many of the same visualization primitives proposed elsewhere by Tversky et al. [56] and Ware [59]. The second step is to understand why such elements carry structural information, so we can come up with a model that explains the effects we have already seen and can help predict effects we have not seen.

We propose such a model based on the implied dynamics of a visualization. This builds on our previous study as well as work in visual cognition that shows that people viewing even an abstract static visual scene tend to simulate physical forces at work in it, and that this simulation gives rise to inferences about scene dynamics. This area of study, "implied dynamics," has shown that people tend to remember objects in static scenes that appear to be moving or to have some force acting on them as farther along in their course of movement than they actually were. Freyd [25] first demonstrated this effect in still photographs. Participants are first shown a photograph of a person in the middle of jumping, then shown a second picture and asked whether they show the same frame of action. People are more likely to misidentify a picture that shows a later moment in the sequence as being the same frame than a picture that shows an earlier moment. Freyd argues that this shows that people mentally simulate the jumping action forward in time, and are therefore more likely to remember the photo they saw as being taken at a later point in the jump. Freyd and colleagues went on to show this effect on more abstract diagrams that implied motion [26], and even in simple cartoon representations that do not directly imply motion but suggest gravity at work [27].

The fact that this effect can be found even in cartoons or abstract diagrams suggests the possibility that it may be at work in infovis as well. If so, this may be a source for some of the ways that people interpret information about structure from a visualization. This possibility recalls theories by Arnheim [6], who argues that the way people interpret the meaning of visual art derives in great part from a sense of the dynamics of composition. That is, certain elements in a scene—be it pictorial or abstract—attract or repel one another because of the viewer's sense of proportion and visual weight, and these implicit forces can be interpreted semantically.

This account of visual perception would seem to explain many of the comments by participants in our study on visual design (Section 6.1.3). This in turn suggests that structural perceptions in infovis may be based on such implied dynamics. In a visualization context, I argue that these dynamics are metaphorically mapped to the target domain of the data to support reasoning about information using a visual representation. This recalls the physical inferences in verbal metaphors, such as the *anger is heat* example described in Section 3.1.

If this model is true, it would explain how people were able to derive structural information from even minor aspects of chart design. In order to test this model, we began to study whether implied forces at work in a visualization can indeed alter a viewer's perception of the position of marks. If implied dynamics of the kind studied by Freyd and others can work on marks in a visualization, this may prove to be a way to predict and study how people perceive visual structure.

7.2 Implied Dynamics and Visualization Use

The first step in establishing implied dynamics as part of the visual metaphor model of visual structure is to test whether these dynamics actually do apply to marks in a visualization. Given the abstract nature of the visual representations in infovis, do people actually see them as objects with forces acting on them? We approached this question initially by designing a study to test whether people simulate gravity when viewing a visualization, as they do to a cartoon drawing of a potted plant in one of Freyd's implied motion experiments [27].

If this is indeed the case, it would also be useful to know whether this effect is constant for all types of marks or whether it is affected by visual and structural properties. For example, some of the comments described in Section 6.1.3 suggest that the apparent weight of an element can affect how much visual momentum it is seen as having. If gravity is indeed a factor and can be predicted, this suggests that implied dynamics may be a valid way to describe how visual relationships are perceived.

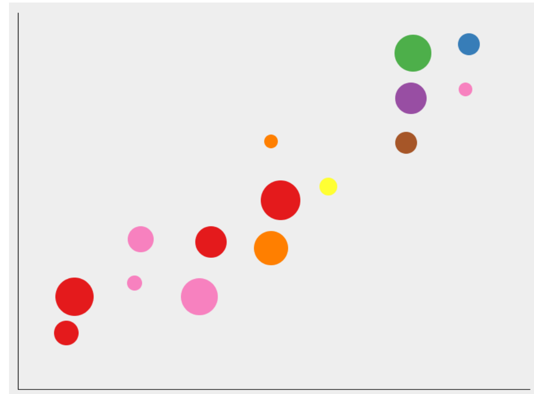
7.2.1 Procedure

To test whether people simulate gravity acting on marks in a visualization, we designed a simple study of visual memory. We recruited 45 participants via Amazon's Mechanical Turk. Participants included 21 females (46.7%) and 24 males. Self-reported age ranged from 18 to 64, with a mean of 30.9.

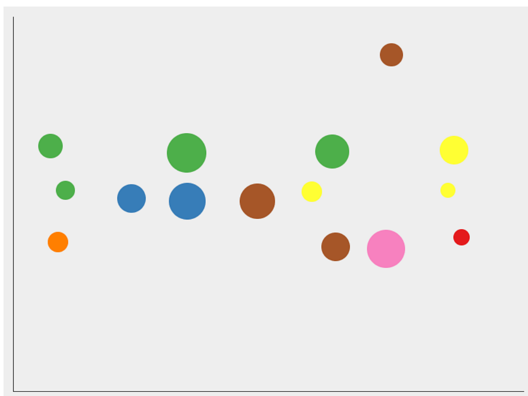
The materials used in this study were a series of "bubble" scatterplot graphs similar to those used in Gapminder [2]. Examples are shown in Figure 15. We used these because they were relatively simple visualizations which allowed for natural variation in the size, color, and layout of visual marks. Each graph contained 15 circles laid out in a semi-random formation. One of these circles was the "target" circle, and participants were told to remember its position and recall it on a blank graph afterward. The overall design was inspired by Freyd's studies of representational momentum and implied dynamics, although we had participants directly report their memory of the mark's position with a mouse click.

We varied several visual factors of the target circle as well as the overall layout of the graph to test whether any perceived gravity was altered by the apparent weight

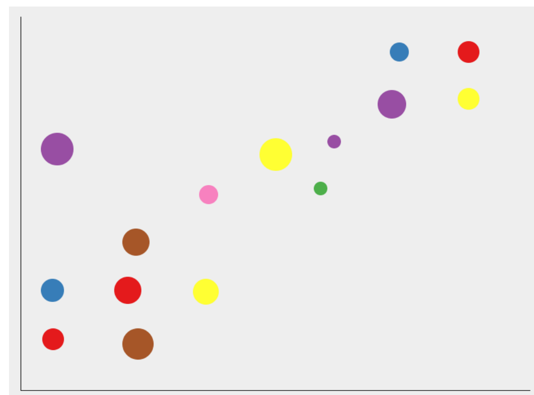
(a) Flat distribution



(b) Linear distribution



(c) Flat distribution, isolated target



(d) Linear distribution, isolated target

Figure 15: Visualizations used to test the effect of gravity on a participant's memory for a visual mark. One of the circles in each of these graphs would flash, and the participant would try to remember its position after the graph vanished.

of the target or its relation to the rest of the distractor points. We altered the color and size of the target, since we hypothesized these factors to have the most direct influence on the target's apparent weight. We also varied its position systematically with a focus on testing whether marks that are higher up in relation to the graph's X-axis are more likely to show a gravity effect. In addition to this, we varied the target's position relative to the distractor points; the distractors were clustered together, and the target was either within this cluster or isolated from it. Finally, this cluster of distractors was either flat or laid out along a roughly linear relationship. These last

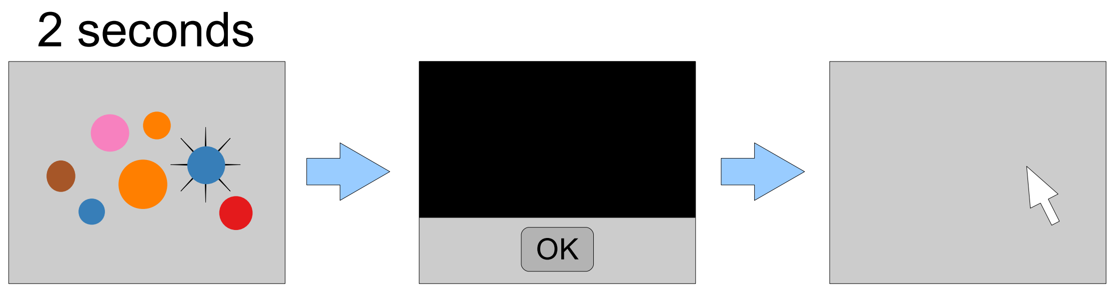


Figure 16: The procedure of a trial in the gravity study. Participants were told to memorize the blinking circle in the first graph, then click on the location of the center of the blinking circle on a blank graph.

two factors were meant to test whether the other marks in a graph, especially those with a strong trendline, have an effect on a target's apparent dynamics that can override or confound the simple gravity effect.

Over the course of the experiment, participants saw 32 of these graphs. In each trial, the target circle initially flashed twice, then the graph remained visible for 2 seconds. The graph was then replaced by a black screen at the bottom of which was an "OK" button that participants had to click to continue. This button was included to prevent participants from leaving their mouse cursor centered over the target circle after the initial graph vanished. The black screen was then replaced by a blank graph in which only the X and Y axes were visible. Participants were told to click their mouse on the location on the blank graph where they remembered seeing the flashing target. An X briefly appeared on the screen where they clicked for feedback. They then continued immediately to the subsequent trial. The design of a single trial is summarized in Figure 16.

We recorded the location of the participant's guess about the target's location as well as their response time. Our primary measures in analyzing the data were the accuracy of their response, measured in Euclidean distance from the actual position of the target, and in particular the amount by which the target was remembered as being lower on the screen than it actually was.

7.2.2 Results

In analyzing our results, we first removed cases with extreme distance error, defined as greater than three standard deviations from the mean error ($M = 35.47$, $S.D. = 73.1$), or greater than 254.8 pixels. This was intended to remove any cases in which a participant was clicking randomly in order to finish the task more quickly, or where she had disregarded or failed to follow the task instructions. This resulted in the dropping of 39 trials, or 2.7% of the total.

Overall, we found a weak but significant overall effect of gravity. We analyzed the measure of vertical error, or the actual position of the target subtracted from its remembered position. A negative vertical error means that the participant remembered the target as being below its actual position, while a positive vertical error means it was remembered as being above its actual position. The average amount by which the response point was below the target point was 1.3 pixels, which is a very minor difference but nonetheless significantly greater than zero ($t(1156) = -3.2$, $p < .01$). Similarly, there was a significant tendency for the response to be to the left of the target point by an average of 1.5 pixels ($t(1156) = -4.0$, $p < .01$). The similarity of the downward and leftward shift suggest the possibility that, rather than a straightforward gravity effect, participants generally remembered points as being closer to the axis lines than they actually were.

Contrary to our hypothesis, we found no effect of target color or size on this downward shift, nor did these object-level properties have a significant effect on the leftward shift. We did, however, find a significant correlation between target height and amount of downward shift ($r(1157) = 0.19$, $p < .01$); that is, circles that were higher up on the graph shifted further down than those closer to the bottom of the screen.

Although this effect is very small, it is worth noting that the findings in the studies of gravity shifts by Freyd et al. [27] also suggested that the amount by which an object shifts downwards in a participant's memory is quite small, with common errors in her experiments falling at 0.09cm to 0.14cm below the actual position. While our experiment used a different testing method, the average amount of downward shift does seem to be similar.

However, the effects of the structural factors we varied suggest a more complex interpretation. As discussed in Section 7.2.1, we varied both the absolute position of the target and its position with respect to the clustered distractor circles. We studied the variable of isolation by splitting the targets into three groups: those which were not isolated from the larger cluster, those which were isolated and above the cluster, and those which were isolated and below the cluster. This effect was studied with a 3x2 repeated measures ANOVA, in which factors were target position and distractor layout (flat or linear) and the dependent variable was the amount of vertical error. The main effect of target position and the interaction of target position by distractor layout both failed the assumption of sphericity, so a Greenhouse-Geisser correction is employed for these tests.

We found a significant main effect of target position, $F(1.47, 43) = 15.06$, $p < .001$,

$\eta^2 = .26$. Pairwise comparisons using a Bonferroni test show that all three cases (isolated above, isolated below, and within the distractor cluster) differed significantly from one another. We found that the downwards vertical shift was most dramatic when the target was isolated above the distractor cluster, and in fact was reversed on average when the target was isolated below the distractor cluster. That is, participants remembered the target as being higher than it actually was when it was positioned underneath the main cluster.

We also found a main effect of distractor layout on vertical error, $F(1, 43) = 6.51$, $p < .05$, $\eta^2 = .13$. That is, when the distractor layout is flat, the target is remembered as being lower than it actually was on average ($M = -4.98$, $S.D. = 1.31$), while the average vertical error is close to zero when the distractors are laid out in a linear trend ($M = .76$, $S.D. = 1.5$).

These findings are further clarified by the significant interaction between target position and distractor layout, $F(1.56, 43) = 6.54$, $p < .01$, $\eta^2 = .13$. This interaction is summarized in Figure 17. The strongest downwards shift is found when the target is isolated above a flat distractor cluster ($M = -17.94$, $S.D. = 3.56$) and the strongest upwards shift, when the target is isolated below a flat cluster ($M = 5.63$, $S.D. = 2.36$). The least amount of average vertical error in either direction is found when the target is within a distractor cluster with a linear trend ($M = -.18$, $S.D. = 2.7$).

Overall, these results suggest that, rather than a straightforward gravity effect pulling marks downwards, there is a tendency to remember the target as being closer to the central mass of distractors than it actually was. This tendency may even out when the distractors are laid out in a linear pattern.

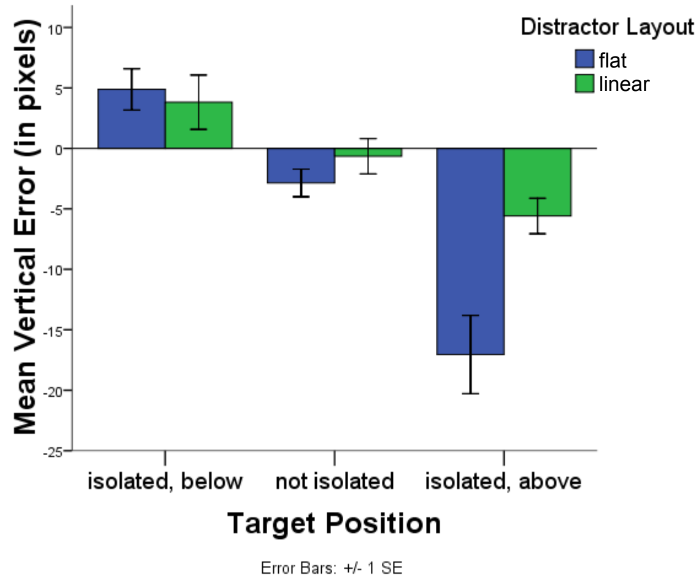


Figure 17: The amount of vertical pixel error in each of the target's three position conditions and the two distractor distribution conditions. A negative value indicates a downward shift, while a positive value indicates an upward shift.

7.2.3 Discussion

While the downward shift we found was not dramatic, it was present and significant, which raises the possibility that people do simulate physical dynamics in visualization scenes. However, there also seems to be a more salient effect of attraction between the target mark and other marks in the graph. It is possible that the target is generally attracted to the mass of distractors, and this effect may be strengthened when the target is above said mass. In this light, and given the presence of an apparent leftward shift as well as a downward shift, it is worth considering that the gravity effect may actually be an attraction to the axis line of the graph, rather than a global physical property.

This apparent attraction is reminiscent of Gestalt grouping principles as well as

Arnheim's theories [6] about visual weight and attractions between visual shapes. It is possible that our results arise not from simulated gravity but from the perceived attraction between marks. Arnheim argues that such attraction is a major factor in the interpretation of composition, and can be altered in various ways by perceptions of visual weight and proportion. Another way to interpret this is that people tend to remember marks as being closer to where they would expect them to be; that is, closer to the average position.

What is particularly interesting in terms of the attempt to include implied dynamics in a model of visual structure is the implication that marks are perceived as being physically closer when people see them as "belonging" together. That is, participants may have felt that the targets belonged in the larger mass, so remembered them as being closer. If this finding that marks that metaphorically belong together are perceived as closer together is a reliable one, it would lend considerable weight to our overall model. Therefore, the next step in this research is to study this attraction effect and what might cause it.

7.3 Factors in Implied Attraction Between Marks

There are a number of standard structural elements that people use to suggest relationships in a visualization, such as connection, grouping outlines, and similarity of color. If the dynamics model works as we have hypothesized, we should expect such structural elements to actually exert attraction between the points they connect, as was hinted in the previous study. Therefore the marks should be seen as physically closer, which may then be interpreted as metaphorically closer.

The hypothesis of the following study is that visual elements that imply conceptual relationships between objects represented by marks in a visualization will also cause those marks to be remembered as closer together. In order to test this hypothesis, we designed a study to find out whether visually linked marks exert an attracting influence on one another.

7.3.1 Procedure

The procedure of this experiment resembled that of the previous one, as described in Section 7.2.1. However, we simplified and focused the design to analyze the extent to which structural elements caused two marks to be remembered as closer together.

As before, participants saw a series of trials in which they were asked to remember the location of a target circle. However, in this case, there were only three circles in each trial, and the participant was asked to remember the location of the center circle (the target). The position of the target varied randomly between trials, but the other two circles were always positioned the same distance from the target along a straight line. This line was positioned either vertically, horizontally, or diagonally. In each trial, one of the two circles on either side of the target was the "attractor" and the other was a distractor (Figure 18).

In each case, the attractor was linked to the target with one of six elements that we hypothesized to suggest a relationship between the two marks (Figure 19). These included three external structural elements: a connecting line, an outline circling the two marks, and a fill behind the marks. There were also two cases where the target and the attractor were linked by similarity, one in which they were the same color

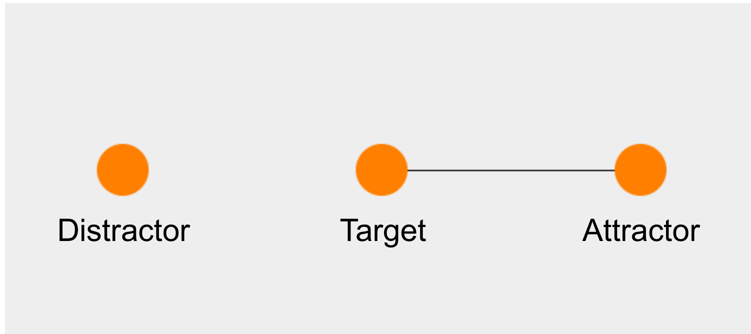


Figure 18: The general layout of the stimuli in the attraction factors study.

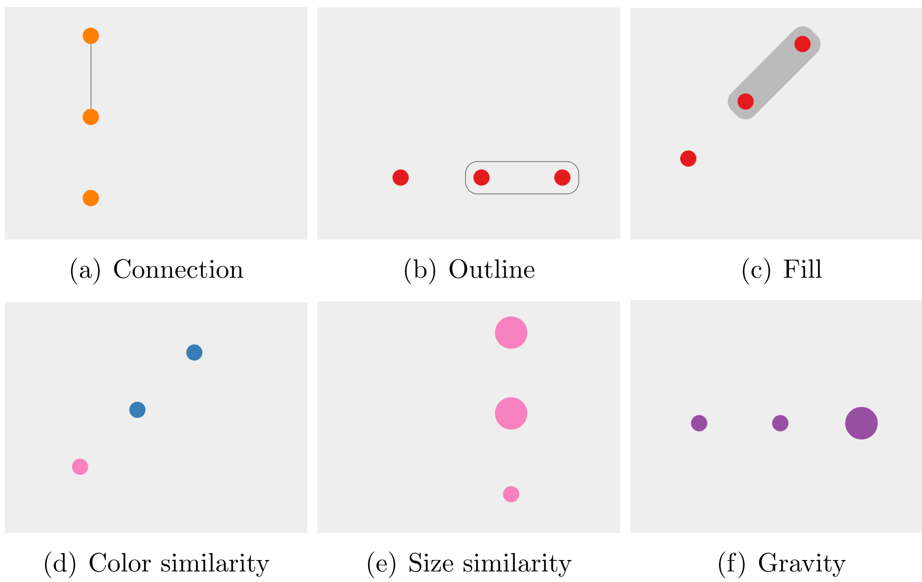


Figure 19: The six elements we hypothesized to cause perceptual attraction between marks.

(and the distractor was a different color) and one in which they were the same size and larger than the distractor. Finally, there was a case in which the attractor was larger than both the target and the distractor. This was meant to test whether an object with an apparently greater "mass" exerted a greater pull on the target, as suggested by Arnheim [6].

Each participant saw each one of these six elements in all six possible orientations; orientation factors included whether the line of marks was vertical, horizontal, or diagonal, as well as which side of the line the attractor was on in each case. Therefore, each participant saw 36 trials. Since the stimulus was much less complex than in the previous study, we showed each image for only 1 second before replacing it with the black screen. In addition, we did not include axis lines as in the previous study. Apart from these differences, the procedure was identical to that in the previous study.

We performed this experiment with 48 participants recruited online through Amazon Mechanical Turk. As before, we measured response time and error measured as Euclidean distance between the participant's mouse click and the center of the target.

7.3.2 Results

As in the previous study, we removed those responses where the overall distance error ($M = 30.59$, $S.D. = 35.89$) was greater than three standard deviations from the mean, which resulted in the removal of 2.6% of the responses. In this experiment, our primary measure was the amount by which the target was remembered as being closer to the attractor than the distractor: that is, the distance between the response point and the center of the distractor minus the distance between the response point and the center of the distractor. We call this metric the attractor shift. Overall we found that the attractor shift ($M = 1.5$, $S.D. = 29.8$) was close to being significantly greater than zero ($t(1411) = 1.9$, $p = .05$), suggesting that in general participants may have remembered the target as being closer to the attractor than it actually was.

Using a repeated measures ANOVA, we found a small but significant main effect of element type, $F(3.99, 46) = 2.59$, $p < .05$, $\eta^2 = .07$, suggesting that the elements we chose exert varying degrees of implied attraction. These results are summarized in

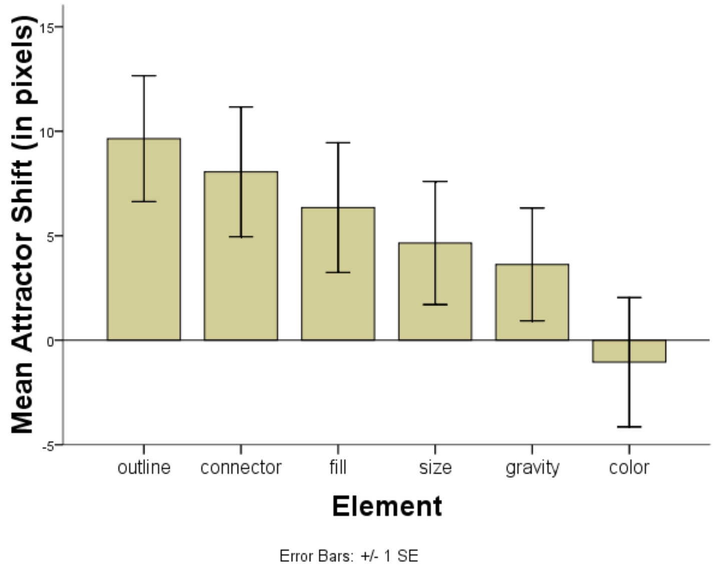


Figure 20: The amount of attractor shift for each of the six relating elements.

Figure 20. The strongest effect was found for outline and connecting line, and color similarity had a slightly negative effect, so that participants remembered circles of the same color as being slightly further apart than they actually were.

In contrast to our previous study (Section 7.2), we did not find that the marks were remembered as being consistently lower than they actually were. In fact, we found an opposite effect in which marks were remembered as shifting upwards ($M = 3.3$, $S.D. = 16.0$), an effect which is significantly greater than zero, $t(1410) = 7.7$,$p < .01$. It is possible that the gravity effect simply vanishes in the absence of a visible X-axis, which may have served as a "ground plane" for viewers. This also supports the speculation that rather than a global gravity effect, the findings in the previous study were largely driven by attraction between the larger mass of distractors and the target.

7.3.3 Discussion

These results lend weight to the idea that implied dynamics between visual marks are metaphorically interpreted as statements about conceptual relationships between data elements. A connecting line is the element we expected to have the strongest effect, and the strong effect of outline is also not surprising. As Tversky et al. [56] and Ware [59] point out, these are some of the strongest cues to relationships among marks in a visualization. The lesser effect of size and background fill, both of which are successfully used to group objects in visualizations, suggests nonetheless that a sense of relationship between marks can arise from factors other than the attraction we found (or perhaps that these are weaker cues to relationship). The nonexistent to negative effect of color similarity is more surprising, and bears further study.

Overall, these results suggest strongly that those elements we see as suggesting conceptual closeness also suggest perceptual closeness. This provides support for the visual metaphor model of visualization structure by showing at least one clear case where a perceptual factor appears to be metaphorically interpreted. The direction of causality is not clear, but this at least shows that there is a relationship between perceived dynamics and information structure.

7.4 Implications of the Implied Dynamics Model

This work shows important initial evidence for the implied dynamics model of how people read visual structure. As outlined in Section 7.1, this model states that a viewer simulates the apparent forces and dynamics at work in a visualization and then metaphorically interprets those dynamics as statements about relationships and patterns within the data. While more work is needed to establish this as a general principle, I have demonstrated that in the specific example of structural elements used to show a relationship between two marks, elements that are used to suggest conceptual closeness can also create a sense of perceptual closeness.

This simple finding has a number of serious implications for infovis theory and practice. Most importantly, it shows that these rather high-level ideas about visual structure and the semantics of design discussed so far can actually have a significant influence on low-level perception. In this case at least, the objection that visual metaphors and structural effects are interesting but irrelevant to practical visualization use does not hold up. Rather, the apparent connection between structural closeness and perceptual closeness suggests that the process of reading data values and the process of understanding visual structure are more tightly linked than the variable encoding model allows.

Given that the context in which a visual mark appears can affect the perception of its position, the variable encoding model clearly does not account for everything that happens in visualization use. However, the results so far do not clearly explain the direction of causality of this effect. That is, do people see marks as being related because they seem perceptually closer, or do marks seem closer because they seem related? It could be that visual illusions and biases that create illusory proximity lead to the conceptual sense of closeness, or that structural elements that imply metaphorical closeness lead to implied attraction. More testing is needed to analyze exactly what is going on in this process.

An interesting implication of these results is that they suggest a possible experimental paradigm for testing the effect of structural elements, at least those meant to imply relationships. Testing whether a given grouping mechanism causes an attractor shift may be an efficient way to validate its usefulness. However, further work is needed to establish a direct correspondence between the degree of attractor shift an element causes and its ability to suggest a conceptual relationship.

This direct connection between perceived closeness and metaphorical closeness, and the support it provides for the implied dynamics model of structural perception, makes the process underlying the visual metaphor model of visual structure considerably more concrete. Given the empirical evidence presented thus far, I believe it is reasonable to treat visual metaphors as a foundation for further visual structure theory. In the following chapter, I will review the implications of these experiments and what this overall framework of visual structure looks like given what we have learned.

CHAPTER 8: TOWARDS A FRAMEWORK FOR INFOVIS STRUCTURE

Taken as a whole, this research tells us a number of interesting things about the importance of visual structure and how people use it to understand data. The ultimate purpose, however, is to put all this information together into a model of how visualization works. I argue that the evidence presented so far supports the visual metaphor framework of visual structure and expands on this by showing that metaphorical interpretations of implied dynamics may underly perceptions of conceptual relationships. Using this new information, we can begin outlining a new perspective of how visualization works based on visual metaphors.

8.1 How Visualization Works

A person viewing a visualization first sees just a collection of shapes. Low level perception produces descriptions of objects and some sense of the overall gist, in terms of how marks are distributed, the dominant colors in the visualization, and so forth. Simple visual grouping can give rise to perceptions of clusters and other simple organizational properties.

At this point, visual structure becomes important. The viewer may use visual decoding to extract object-level information, but only for a limited number of marks at any given time. At a higher level, the viewer is beginning to get a sense of the structure of the visualization. This happens initially through simple simulations of the dynamics acting on the image. The field of vaguely organized visual marks becomes a collection of objects acting on one another, through forces such as attraction, repulsion, support, and connection.

Once these dynamics have been perceived, the viewer can metaphorically transform these simulations into inferences about data relationships. For example, an apparent attraction between two marks suggests to the viewer that the objects they represent have something in common, while an apparent repulsion may suggest that those objects have opposing goals. At this point, knowledge of the semantic domain of the data may also come into play to constrain or suggest certain metaphorical interpretations, as may the user's own preconceptions and expectations.

Once such inferences start being generated, the user checks them by reference to individual data points, which is when the efficiency and accuracy of variable decoding becomes an important part of the process. This object-level data is interpreted in context, however. The viewer will already be forming hypotheses about the data based on the visual structure, and her understanding of possible data relationships will be powerfully constrained by the mental model she has begun to form.

The process of using a visualization, then, is a cycle of forming expectations about structure, forming hypotheses based on that structure, and checking to see if low-level data conforms to those hypotheses. The results of those checks will be incorporated into the ongoing mental model of the data, which will then lead to new hypotheses and expectations. Structure and data interact constantly in the user's attempt to understand a problem or phenomenon, and the strength of a good visualization lies in its ability to model its data in a way that supports this interaction meaningfully.

8.2 Implications for Design

The tools presented here for describing and analyzing visual structure in terms of visual metaphors can be applied in various ways to improving and better understanding infovis design.

Explicitly note the metaphors being used. Designers that think about visual metaphors currently tend to focus on high-level structural metaphors, such as IN-SPIRE's *ideas are points in a landscape* metaphor [33] or ThemeRiver's *the flow of events is the flow of water* [32]. But all visualizations use ontological and orientational metaphors, even if designers of less obviously metaphorical visualizations do not explicitly think about which low-level metaphors they use. The strong effect of metaphorical conflicts found in Chapter 5, which were found in visualizations that are not thought of as being highly metaphorical, demonstrate how dangerous it can be to be unaware of the implicit metaphors in your visualization. Designers should plan metaphors from the beginning so these conflicts can be avoided or at least predicted.

Control structure by analyzing dynamics. The findings in Chapter 6 show how broadly the implied dynamics of even very simple charts can affect a user's perception of information. Specifically, they show the importance of apparent support, stability, and connectivity in implying how pieces in a chart relate to one another. By viewing a visualization as a collection of physical objects and analyzing how those objects appear to enact forces on one another, a designer can control these effects and avoid unintended interpretations. Additionally, dynamics can be used to suggest possibilities and constraints for both interaction and reasoning, similarly to the use of affordances in human-computer interaction.

Novel visual metaphors may be harder to understand at first. Many of our results showed greater efficiency and ease of understanding with traditional visualizations like a node-link diagram (Section 5.1) or pie chart (Section 6.1) than with more novel visualizations. This is likely analogous to the idea of conventional and novel metaphors in language; conventional visual metaphors require less translation, and therefore are faster to understand. However, they also lead to less elaboration and inference-making, which may or may not be a good thing. The extra cognitive effort required by novel metaphors may make them more engaging for users, which may be the goal in some cases.

Consistent metaphors across a visualization system will lead to more efficient use. Our findings show that switching between metaphors carries a cognitive cost for many users (Chapter 5). If an infovis system contains multiview displays or mixes text and visualization, any conflicts in metaphor among the different parts of the system can slow users down. However, this may not always be a bad thing; it is possible that such structural switching can cause users to think harder about a problem. Designers should consider whether the goal is efficient information extraction or complex exploration.

None of this may work the same way for expert users. One of our most striking findings is that users with a greater tendency to flexible thinking were less likely to have difficulty switching between conflicting metaphors (Section 5.2). This suggests that users with innately higher spatial ability, and perhaps as well those with greater knowledge of a particular domain, may be less affected by visual structure than others. If you are designing for such a group, you may have more freedom in the kind of designs you can try out, whereas designs for a wider audience should avoid structural conflicts to a greater degree.

8.3 Implications for Evaluation

A second clear area in which this model can improve visualization practice is in evaluation and user studies of visualization systems. Our findings show that not taking structure into account can lead to unintended biases, and more fundamentally, that the usual paradigm of evaluation is not capturing all the differences between methods. Two methods may have the same efficiency, but imply drastically different things about data patterns and relationships. Infovis evaluation should make the attempt to understand these differences.

Find out what inferences a visualization leads people to in a task-free setting. The findings in Section 6.1 suggest that much can be learned about how a visualization structures information by having a user generate hypotheses about the data in free-form exploration. This is similar to proposals for insight-based evaluation in infovis [49], but I would argue that such evaluations should be focused on the *kinds* of insights a user has, not just the number. Do different users come up with the same inferences? How confident are they in these inferences? Are their findings consistent with real data properties, or are they suggested by unintended effects in the visualization dynamics? Answering these questions would be key to understanding what kind of reasoning a visualization system supports.

Find out how design-sensitive the visualization is. In Section 6.1, we describe the finding that semantic responses to certain chart types, such as pie charts, seemed to be less susceptible to changes in design than others, such as bubble charts. Knowing how much flexibility a given visualization offers in terms of design would be a very useful thing to know, and could be tested by performing user studies similar to those in our study.

Make sure people understand what metaphors are being used. One of the benefits of designing a visualization around an explicit set of visual metaphors is that you can then more easily measure whether your users understand the visual structure of the system in the way you intended. By prompting the user to explain the visualization as they understand it, you can check that the verbal metaphors they use match the ones you want them to perceive While it is never easy to establish what a user's mental model entails, it should at least be possible to find glaring inconsistencies that may be confounding evaluation results, as found in Chapter 5.

CHAPTER 9: CONCLUSION

I argue that visual metaphors make a powerful and useful framework for visualization, and that the empirical evidence I have presented strongly supports its validity in practice. That said, this is meant to be a foundation for a body of theory, not a final pronouncement on that theory. A number of open questions remain. Most importantly, this work has not yet addressed interaction in any way. The work presented here deals only with static visualizations, which was the result of a conscious decision to limit the complexity of the experiments. But in practice, infovis research almost never deals with purely static representations. Further, given the relationship between the model I present and concepts from human-computer interaction such as affordances and mental models, it seems likely that visual metaphors and implied dynamics could add a great deal to our understanding of how people learn to interact with a visualization and how that interaction changes her perception of data.

A related gap in this theory so far is how a task affects a user's understanding of visual structure. When a user has a specific problem she wants to solve using a visualization, she will likely have different demands on, and expectations of, its information structure. Similarly, the semantics of the data being visualized likely affect the perception of structure in subtle and dramatic ways (although Zacks and Tversky's study of bars and lines [63] suggests that visual structure can override data semantics in some cases). In general, a better understanding of visual structure will require understanding of the context in which a user approaches a visualization.

While there are a number of practical questions that still need to be answered, this model at least makes it possible to ask these questions, and gives us some guidance on how to go about answering them. Not only do I feel that this theoretical direction is relatively well-supported at this point, but I believe that incorporating some understanding of visual structure into our view of visualization is absolutely necessary if we are to make sense of the role of infovis in the greater world.

A look at the way visualization gets used in the real world shows the weakness of the variable encoding model on its own. In a typical newspaper visualization, such as those found in *The New York Times*, you see not only the visual representation but also a number of pointers to interesting parts of the graph and text to explain why they're interesting. In a typical Powerpoint presentation—probably the most common place for people to see visualization of information—graphs are accompanied by captions, not to mention a speaker who explains the context and of what is presented and point out what the audience should be observing. Even when you move to more complicated, exploratory tasks with a visualization system, you don't necessarily find users needing a greater ability to read data values. Indeed, understanding broad patterns and data dynamics is probably more important than any individual data value in a high-level analytical application.

If real-life use of visualization usually includes some kind of non-visual guidance, why do people use the visualization at all? Why not just write an article or put your points in a bulleted list? If infovis is just about presenting data to people in a form that they can efficiently decode, these common uses of infovis seem pointless. And yet people clearly believe that the visualization adds something beyond a presentation of data points.

For all our focus in the infovis community about increasing the bandwidth of information and the efficiency of data retrieval, the point of infovis in practice is not to make it easier to read object-level data. Rather, what all these real-world uses of infovis tell us is that the real strength of visualization is in providing a context and a structure for the data values that are read from it. The purpose of infovis is to get a sense of what information means, which means understanding the structure of that information. When you understand the structure, the details make more sense, and you can begin to see relationships and patterns.

And yet the way we design, evaluate, and theorize about infovis continues to focus on the details without stepping back to look at the structure. In many ways, we still build and evaluate visualizations from the ground up, by analyzing data mappings and how well they can be decoded. Until we adopt a model of visualization that takes structure into account, we will forever be missing the forest for the trees. I believe that the evidence presented in this thesis provides strong support for visual metaphors as a sound basis for the development of such a model.

The ultimate point of such work is that the first thing a real user sees in a visualization is not an array of data points; the first thing a user sees in a visualization is the visualization. And, as with the millennia of visual image-making that preceded it, that visualization tells you something about the structure of the information it contains. Making sense of this information in the context of a visual structure is how visualization works.

REFERENCES

[1] Amazon Mechanical Turk, http://aws.amazon.com/mturk/. Retrieved 2-11-10.

[2] Gapminder, http://www.gapminder.org. Retrieved 2-11-10.

[3] Tableau, http://www.tableausoftware.com/. Retrieved 2-11-10.

[4] Agutter, J., Syroid, N., Drews, F., Westenskow, D., Bermudez, J., and Strayer, D. Graphic data display for cardiovascular system. In *IEEE Information Visualization* (2001), IEEE CS Press, pp. 163–166.

[5] Andrews, K., and Kasanicka, J. A comparative study of four hierarchy browsers using the hierarchical visualisation testing environment (HVTE). In *Proceedings International Conference Information Visualization (IV)* (2007), IEEE CS Press, pp. 81–86.

[6] Arnheim, R. *The Power of the Center: A Study of Composition in the Visual Arts*. UC Press, 1988.

[7] Barlow, T., and Neville, P. A comparison of 2-D visualizations of hierarchies. In *Proceedings Information Visualization* (2001), IEEE CS Press, pp. 131–138.

[8] Bauer, M. I., and Johnson-Laird, P. N. How diagrams can improve reasoning. *Psychological Science 4*, 6 (1993), 372–378.

[9] Bertin, J. *Semiology of Graphics*. Univ. of Wisconsin Press, 1967.

[10] Booth, P. A. *An Introduction to Human-Computer Interaction*. Psychology Press, 1989.

[11] Bowdle, B. F., and Gentner, D. The career of metaphor. *Psychological Review 112*, 1 (2005), 193–216.

[12] Brewer, C., and Harrower, M. ColorBrewer, http://www.colorbrewer.org/, 2002. Retrieved 12-4-2008.

[13] Card, S. K., and Mackinlay, J. The structure of the information visualization design space. In *Proceedings Information Visualization* (1997), IEEE CS Press, pp. 92–99.

[14] Chen, C. Top 10 unsolved information visualization problems. *IEEE Computer Graphics and Applications 25*, 4 (2005), 12–16.

[15] Chen, C., and Yu, Y. Empirical studies of information visualization: A meta-analysis. *International Journal of Human-Computer Studies 53* (2000), 851–866.

[16] Chi, E. H. A taxonomy of visualization techniques using the data state reference model. In *Proceedings Information Visualization* (2000), IEEE CS Press, pp. 69–76.

[17] Cleveland, W. S., and McGill, R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association 79*, 387 (1984), 531–554.

[18] Cooke, N. J., and Bartha, M. C. An empirical investigation of psychological metaphor. *Metaphor and Symbol 7*, 3 & 4 (1992), 215–235.

[19] Cox, D. Metaphoric mappings: The art of visualization. In *Aesthetic Computing*, P. A. Fishwick, Ed. MIT Press, 2006, ch. 6, pp. 89–114.

[20] Donnellan, M. B., Oswald, F. L., Baird, B. M., and Lucas, R. E. The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment 18*, 2 (2006), 192–203.

[21] Eckstrom, R. B., French, J. W., Harman, H. H., and Derman, D. *Kit of factor-referenced cognitive tests*. Educational Testing Service, Princeton, NJ, 1976.

[22] Elting, L., Martin, C., Cantor, S., and Rubenstein, E. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. *BMJ 318* (1999), 1527–1531.

[23] Ericson, M. Visualizing data for the masses: Information graphics at *The New York Times*. Keynote presentation at IEEE InfoVis, 2008.

[24] Fekete, J.-D. The InfoVis Toolkit. In *Proceedings Information Visualization* (2004), IEEE CS Press, pp. 167–174.

[25] Freyd, J. The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics 33*, 6 (1983), 575–581.

[26] Freyd, J. Dynamic mental representations. *Psychological Review 94*, 4 (1987), 427–438.

[27] Freyd, J., Pantzer, T., and Cheng, J. Representing statics as forces in equilibrium. *Journal of Experimental Psychology 117* (1988), 395–407.

[28] Gattis, M., and Holyoak, K. J. Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition 22*, 1 (1996), 231–239.

[29] Gibson, J. J. *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979.

[30] Glenberg, A. M., and Langston, W. E. Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language 31*, 2 (1992), 129–151.

[31] Halpern, D. F. *Sex Differences in Cognitive Abilities*. Lawrence Erlbaum Associates, 2000.

[32] Havre, S., Hetzler, B., and Nowell, L. ThemeRiver: Visualizing theme changes over time. In *Proceedings Information Visualization* (2000), IEEE CS Press, pp. 115–123.

[33] Hetzler, E., and Turner, A. Analysis experiences using information visualization. *IEEE Computer Graphics and Applications 24*, 5 (2004), 22–26.

[34] Horn, W., Popow, C., and Unterasinger, L. Metaphor graphics to visualize ICU data over time. In *Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-98)* (1998).

[35] Kieras, D. E., and Bovair, S. The role of mental models in learning to operate a device. *Cognitive Science 8* (1984), 255–273.

[36] Kittur, A., Chi, E. H., and Suh, B. Crowdsourcing user studies with Mechanical Turk. In *SIGCHI* (2008), pp. 453–456.

[37] Kobsa, A. User experiments with tree visualization systems. In *Proceedings Information Visualization* (2004), IEEE CS Press, pp. 9–16.

[38] Kosara, R., and Ziemkiewicz, C. Do Mechanical Turks dream of square pie charts? In *BELIV Workshop* (2010).

[39] Kövecses, Z. *Metaphor: A Practical Introduction*. Oxford University Press, 2002.

[40] Lakoff, G., and Johnson, M. *Metaphors We Live By*. University of Chicago Press, 1980.

[41] Larkin, J., and Simon, H. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science 11* (1987), 65–99.

[42] Mackinlay, J. Automating the design of graphical presentations of relational information. *Transactions on Graphics 5*, 2 (1986), 110–141.

[43] Mackinlay, J. D., Hanrahan, P., and Stolte, C. Show Me: Automatic presentation for visual analysis. *IEEE TVCG 13*, 6 (2007), 1137–1144.

[44] McQuarrie, E., and Mick, D. Visual rhetoric in advertising: Text-interpretive, experimental, and reader-response analyses. *Journal of Consumer Research 26*, 1 (1999), 37–54.

[45] McQuarrie, E., and Phillips, B. Indirect persuasion in advertising how consumers process metaphors presented in pictures and words. *Journal of Advertising 34*, 2 (2005), 7–20.

[46] Monroe, M. The apparent weight of color and correlated phenomena. *The American Journal of Psychology 36*, 2 (1925), 192–206.

[47] Norman, D. Affordance, conventions, and design. *Interactions 6*, 3 (1999), 38–43.

[48] Norman, D. A. *The Design of Everyday Things*. Doubleday Business, 1990.

[49] North, C. Toward measuring visualization insight. *IEEE Computer Graphics and Applications 26*, 3 (2006), 6–9.

[50] Shneiderman, B. Tree visualization with tree-maps: 2-D space-filling approach. *ACM Transactions on Graphics 11* (1992), 92–99.

[51] Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualization. In *Proceedings IEEE Symposium on Visual Languages* (1996), IEEE CS Press, pp. 336–343.

[52] Stasko, J., Catrambone, R., Guzdial, M., and McDonald, K. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies 53* (2000), 663–694.

[53] Stasko, J., Görg, C., and Liu, Z. Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization 7*, 7 (2008), 118–132.

[54] Stenning, K., and Oberlander, J. A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science 19*, 1 (1995), 97–140.

[55] Tversky, A., and Kahneman, D. The framing of decisions and the psychology of choice. *Science 211*, 4481 (1981), 453–438.

[56] Tversky, B., Zacks, J., Lee, P., and Heiser, J. Lines, blobs, crosses and arrows: Diagrammatic communication with schematic figures. In *Theory and Application of Diagrams* (2000), Springer-Verlag, pp. 221–230.

[57] van Ham, F., and van Wijk, J. J. Beamtrees: Compact visualization of large hierarchies. In *Proceedings Information Visualization* (2002), IEEE CS Press, pp. 93–100.

[58] Viégas, F. B., Wattenberg, M., van Ham, F., Kriss, J., and McKeon, M. ManyEyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1121–1128.

[59] Ware, C. *Visual Thinking for Design.* Morgan Kaufmann, 2008.

[60] Weber, M., Alexa, M., and Müller, W. Visualizating time-series on spirals. In *IEEE Information Visualization* (2001), pp. 7–13.

[61] Wilkinson, L. *The Grammar of Graphics*, 1st ed. Springer, 1999.

[62] Wright, W., Schroh, D., Proulx, P., Skaburskis, A., and Cort, B. The Sandbox for analysis: concepts and methods. In *ACM SIGCHI* (2006), ACM Press.

[63] Zacks, J., and Tversky, B. Bars and lines: A study of graphic communication. *Memory & Cognition 27*, 6 (1999), 1073–1079.

[64] Ziemkiewicz, C., and Kosara, R. The shaping of information by visual metaphors. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1269–1276.

[65] Ziemkiewicz, C., and Kosara, R. Preconceptions and individual differences in understanding visual metaphors. *Computer Graphics Forum, Proceedings Eurographics/IEEE Symposium on Visualization 28*, 3 (2009), 911–918.

[66] Ziemkiewicz, C., and Kosara, R. Implied dynamics in information visualization. In *Advanced Visual Interfaces* (2010).