Eurographics/ IEEE-VGTC Symposium on Visualization 2009
H.-C. Hege, I. Hotz, and T. Munzner
(Guest Editors)

Volume 28 (2009), Number 3

Preconceptions and Individual Differences in Understanding Visual Metaphors

Caroline Ziemkiewicz[1] and Robert Kosara[1]

[1]Charlotte Visualization Center, UNC Charlotte, USA

Abstract

Understanding information visualization is more than a matter of reading a series of data values; it is also a matter of incorporating a visual structure into one's own thinking about a problem. We have proposed visual metaphors as a framework for understanding high-level visual structure and its effect on visualization use. Although there is some evidence that visual metaphors can affect visualization use, the nature of this effect is still ambiguous. We propose that a user's preconceived metaphors for data and other individual differences play an important role in her ability to think in a variety of visual metaphors, and subsequently in her ability to use a visualization. We test this hypothesis by conducting a study in which a participant's preconceptions and thinking style were compared with the degree to which she is affected by conflicting metaphors in a visualization and its task questions. The results show that metaphor compatibility has a significant effect on accuracy, but that factors such as spatial ability and personality can lessen this effect. We also find a complex influence of self-reported metaphor preference on performance. These findings shed light on how people use visual metaphors to understand a visualization.

Categories and Subject Descriptors (according to ACM CCS): Information Interfaces and Presentation (HCI) [H.5.m]: Miscellaneous—

1. Introduction

As information visualization matures, there is increasing interest in developing a theory of visualization. While this line of research has historically focused on formalization of data encodings and developing perceptual guidelines (e.g., Bertin [Ber67]), we propose that using a visualization is not simply a matter of reading data properties encoded in visual form; rather, there are complex cognitive activities at work that are influenced by visual structure. Using information visualization is a process of adapting information from an external visual structure to one's internal conceptual structures, and vice versa. This process can be affected in subtle ways by the conceptual structure of a given information visualization, as well as by the user's individual thinking style and preconceptions.

We have adopted a framework of visual metaphor to address these questions of conceptual structure in information visualization. We see visual metaphors as analogous to verbal metaphors in their ability to shape information and make sense of abstract patterns and relationships. This perspective draws on the idea of mental models and design metaphors from human-computer interaction, and allows us to focus on how the visual structure of an information visualization affects how a user reasons with data.

Equally important is the question of how a user's own conceptual structures affect her use of a visualization. Work in cognitive science [LJ80] suggests that the conceptual metaphors that permeate language are also used to structure abstract thought of all kinds. This makes metaphor a potential framework to analyze not only the structure of a visualization, but a user's preconceived ideas about information.

In this work, we study how users' preconceived structures of information and other individual differences affect their ability to understand visual metaphors in an information visualization. We use tree visualizations as a testbed because of their clear spatial structure and association with similar verbal metaphors. For example, a traditional node-link diagram bears similarities to a metaphor *a hierarchy is a series of levels*, and lends itself to statements about nodes in the tree being *above* or *below* each other. A treemap, on the other hand, embodies the metaphor *a hierarchy is a set of nested containers*, which is reflected in statements about nodes in a tree *containing* or being *inside* one another. We hypothesize that, when using a visualization, a user will internalize its metaphors so that questions asked in a compatible verbal metaphor will be easier to answer, while those which are incompatible (such as a *levels* question asked about data in a treemap) will prove more difficult.

We study the degree to which a user has internalized a visual metaphor by measuring the extent to which these sets of visual and verbal metaphors interfere with each other in simple visualization tasks. This work builds upon the findings of a previous study [ZK08] which showed that in some cases, users showed a faster response time with questions in a compatible metaphor. However, the ambiguity of the results in this earlier research led us to investigate the role of individual differences in the extent to which a user adopts a visual metaphor to her own thinking.

To this end, we examine how factors such as gender, spatial reasoning ability, and self-reported metaphor preference increase or decrease this effect. We study an online population of users to find whether differences in these factors can predict the ease with which a user can answer questions about data in an incompatible verbal metaphor to the visual metaphor of the visualization she is using. These findings will ultimately build toward a model of how we use visual structure to understand and reason with data.

2. Related Work

Our decision to use visual metaphor as a framework to consider the higher-level cognitive effects of individual differences is motivated by work in cognitive science demonstrating that metaphors are used to shape information on a conceptual as well as a linguistic level. This theory is most fully described by Lakoff and Johnson [LJ80], who argue that metaphors in language reflect complex conceptual systems that help us understand and analyze complex abstract topics, such as time or emotion, by structural analogy to simpler and more concrete topic domains. These conceptual metaphors can be implemented verbally, but they can also be reflected in the visual presentation of information, a phenomenon well-studied in the realm of advertising [MM99]. Lakoff and Johnson further distinguish structural metaphors (e.g., *time is a journey*), which involve multiple related mappings between two conceptual domains, from ontological metaphors (e.g., *the mind is a container*) and orientational metaphors (e.g., *more is up*), which involve single mappings between an abstract concept and a type of object or a spatial direction. While visualizations occasionally strive to embody full structural metaphors [HPU98], these ontological and orientational metaphors more commonly underly the visual metaphors of information visualization.

Further work in cognitive science has established the effect of conceptual metaphors on thinking. Gentner and Gentner [GG83] find that presenting users with differing metaphors for electricity can influence their responses on comprehension questions, suggesting that metaphors are used in reasoning with abstract information. Tying conceptual metaphors more directly to visual metaphors, Gibbs and O'Brien [GO90] show that the imagery associated with the verbal metaphors in idioms is remarkably constant across different people and even across different verbal metaphors with the same underlying meaning. They go on to argue that these mental images are constrained by their associated ontological conceptual metaphors. This suggests not only that conceptual metaphors can constrain thinking, but also that visual and verbal metaphors are fundamentally connected.

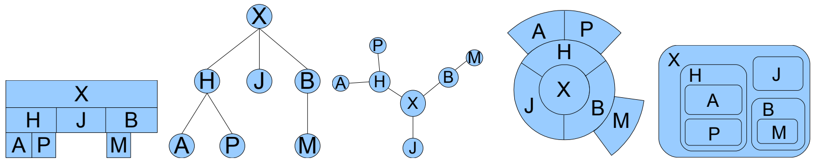
Taken together, these findings support our theory that visual metaphors in a visualization can embody conceptual metaphors which provide constraints and logical entailments when working with visualized information. A previous study by the authors [ZK08] showed that visual metaphors and verbal metaphors can interact to affect a user's response time in simple tasks, and that this compatibility effect is associated with higher overall accuracy on such tasks. While this provided initial evidence that internalizing visual metaphors is an important part of visualization use, the variability among participants suggests that a better understanding of what factors drive this internalization is needed.

While there has been work on the potential role of individual differences in visualization use, for the most part it has focused on low-level perceptual tasks. Conati and Maclaren [CM08] find that a user's perceptual speed predicts whether a star graph or heatmap will be most effective. They did not find an effect of other cognitive factors, including the spatial ability factor we study in the current work. Similarly, Allen [All00] finds a role for perceptual speed and spatial scanning ability in search performance. Taking a more cognitive perspective, Chen [Che00] found no effect of spatial ability on participants' search performance in a visualization of paper citation links, but did find that users with high spatial ability rated their tasks higher on familiarity. Chen goes on to suggest that users with varying cognitive abilities may be more or less likely to impose certain kinds of mental models on data, a central question of this work.

3. Experiment

To test the role of individual differences in internalizing visual metaphors, we conducted a study 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.

We hypothesize that participants will be faster and more accurate on the compatible blocks, and that this effect will be stronger for more difficult tasks. This effect will be influenced by the Openness dimension of a participant's personality and their spatial ability, since users with high scores



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

Figure 1: *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.*

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.

3.1. Participants

63 participants were recruited from Amazon's Mechanical Turk web service [Ama], which allows the assignment of simple tasks to a large population of users online. 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$).

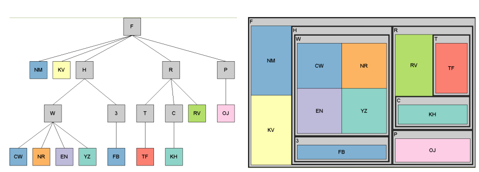
3.2. Materials

Participants initially filled out three scales in web forms meant to measure individual differences that may affect their performance in the study. Personality differences were assessed using the Mini-IPIP Big Five personality scale [DOBL06]. This twenty-question scale rates participants on five major personality dimensions: Openness, Conscientiousness, Extraversion, 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* [EFHD76]. 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." In this paper, we will 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 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 1).

During the test portion, participants answered simple



(a) Node-link condition (b) Treemap condition

Figure 2: *The visualizations used in the study.*

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?

Table 1: *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.*

questions about a hierarchy (described as the files on a computer hard drive) visualized as either a treemap or a node-link diagram (Figure 2). These visualizations were built with the InfoVis toolkit [Fek04] and used a categorical color scheme from ColorBrewer [BH02]. Participants were told that the colors mapped to file types, although this information was not used in any of the task questions. The visualizations were static and non-interactive. 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.

3.3. Procedure

After responding to the three surveys, participants began the main study portion, 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 were worded in either a *containers* or a *levels* metaphor (Table 1).

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 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.

The design of this study was influenced by lessons learned in our previous study on the effect of visual and verbal metaphors [ZK08]. In that work, 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.

4. Results

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.

4.1. Overall performance

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

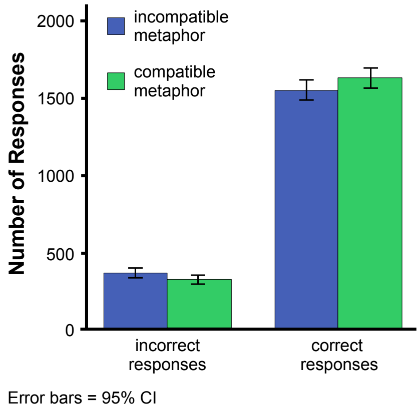


Figure 3: *(a) There was an overall significant effect of question compatibility on response correctness. Participants were generally more likely to answer compatible questions correctly and incompatible questions incorrectly. However, there was no such significant difference for participants with above-average spatial ability, participants who scored highly on the personality dimension of Openness, or participants who reported no verbal metaphor preference.*

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.

4.2. 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 3). 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, as discussed in Section 3.3, 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 [EFHD76]), 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 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 [DOBL06]) 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.

4.3. 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, $F(2,3791) = 5.24, p < .01$.

4.4. 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 3.2. 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

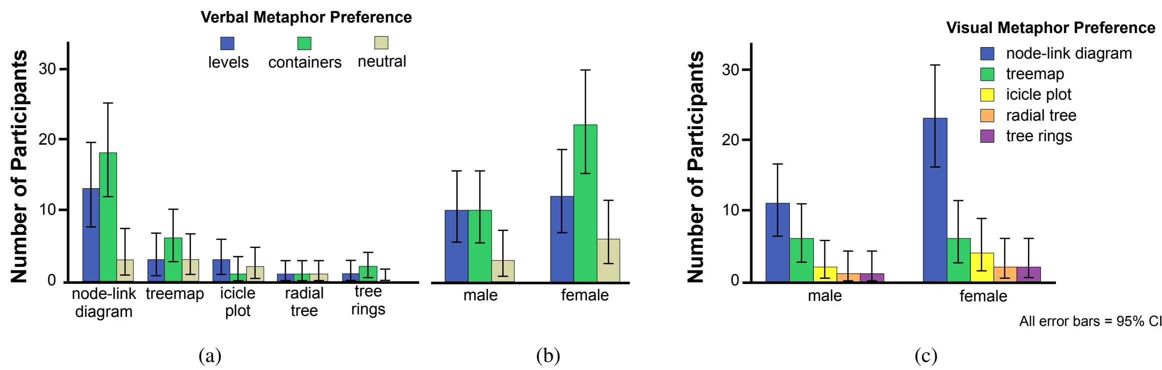


Figure 4: *(a) While these differences are not significant, patterns of self-reported verbal metaphor preference can be seen among participants who ranked a given visual metaphor (Figure 1) as the best depiction of a hierarchy. Contrary to our hypothesis, users who ranked the node-link diagram highest as well as those who ranked the treemap highest tend to prefer* containers *verbal metaphors. Those who ranked an icicle plot highest mostly preferred* levels *verbal metaphors, although this is a small group. Non-significant but notable gender differences also emerged in self-reported metaphor preference. (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.*

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 4(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 participants' preferring 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. Further study is needed to reliably determine the conceptual metaphors of a given visualization.

There were some gender-related patterns in these measures of self-reported preference. Women were more likely than men to prefer *containers* metaphors (Figure 4(b)), although this effect is not significant. However, women also non-significantly tended to rank the treemap lower in their preferred visual metaphors (Figure 4(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.

4.5. 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 5). 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

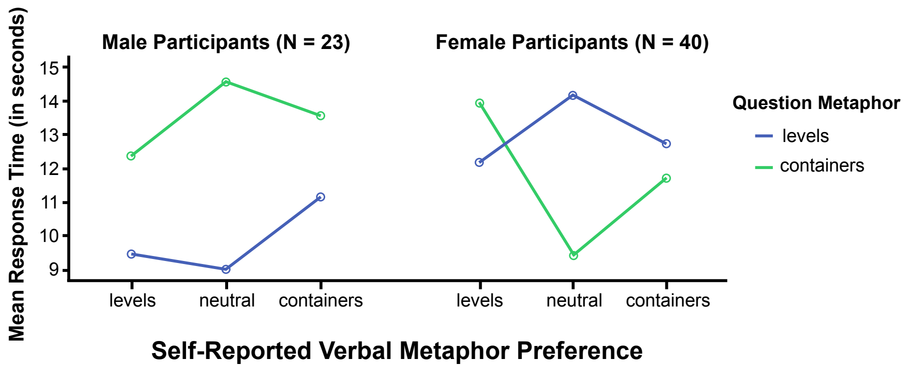


Figure 5: *While we expected users to respond faster to questions in their self-reported preferred verbal metaphor, this pattern only emerged for women. Men responded significantly faster to levels questions no matter what their self-reported metaphor preference. Women who did not prefer one verbal metaphor over another tended to respond faster to containers questions.*

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.

4.6. 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.

5. Discussion

Our results suggest a number of expected and unexpected interpretations of how individual differences contribute to the role of visual metaphors in using a visualization. While metaphor compatibility does influence how well a user can perform on visualization tasks, this influence can be lessened in cases where the user is more inclined to spatial or imaginative thinking, or where the user does not have a strong prior metaphor preference. We further discuss these effects in the areas of overall metaphor compatibility, gender effects, and the relationship between preference and performance.

5.1. Metaphor compatibility

The effect of metaphor compatibility on reading time and correctness suggests that verbal and visual metaphors are significant factors in a user's comprehension of a visualization task. We posit that during the time a user is working with a given visualization, she will try to adapt her thinking to its metaphorical structure. Questions that do not easily fit into that structure will be more difficult to understand, which manifests itself in the time she takes to read the question as well as her ability to answer it correctly.

However, this work also makes it clear that this adapting to a visual metaphor does not always take place. 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.

5.2. Gender

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 [Hal00]. These findings are a clear call to ensure balanced gender distribution in visualization evaluations, particularly ones which use novel visual metaphors that may interfere with a user's preconceptions.

5.3. Preconceptions

The interaction between stated metaphor preference and task performance is not as clear-cut as we expected. Users who prefer a visual metaphor do not always perform better with a visualization that uses that metaphor, suggesting that preconceived conceptual metaphors do not strongly influence a user's ability to adopt a novel visual metaphor.

Although the effect of verbal metaphor preference is similarly ambiguous, it is of interest that participants who show no preference for either of the verbal metaphors are more likely to answer compatible and incompatible questions with equal accuracy. This suggests that, although the visual metaphor seems to dominate while one is using a visualization, it is through verbal metaphors that a participant's preconceptions are more clearly discovered.

6. Conclusion and Future Work

Our work suggests the need to take individual differences among users into account when designing and evaluating visualization systems. We provide evidence that these differences go beyond low-level perceptual effects to influence a user's conceptual structures of information.

These findings imply a need to better understand the needs of individual users when designing visualization methods, not just at the level of visual properties such as size and color, but also when crafting the visual metaphors that drive a visualization's overall design. The mismatch between initial preference and performance suggests that this isn't just a matter of designing to the user's preconceived metaphors, although that may play a part in how the user will formulate questions and ideas about the data.

The differences in gender, spatial ability, and personality should also be considered when evaluating systems. An unbalanced or insufficiently large participant population may skew the results of even high-level tasks. At the same time, these differences may help to account for some of the inconsistencies among existing visualization evaluations.

More work is needed to place the results of this research in a broader context. In particular, a study of the role of verbal ability on the effect of metaphor compatibility may clarify the gender differences we found. It would also be useful to examine how verbal and visual metaphor preferences behave for other types of data than tree visualizations. While tree visualization has the advantage of being highly structured and therefore associated with strong spatial metaphors, these findings must be generalized.

Although we did not have enough non-native English speakers in our population to study the role of language in metaphor effects, a more focused study may be able to shed light on this factor. For example, cognitive science studies have found differences between native English and native Chinese speakers in spatial metaphors for time; exploring the effect of this difference on time-series visualizations may uncover another significant area of individual difference.

Ultimately, we wish to fit this work into a broader model of how people use the visual metaphors of a visualization to think about and reason with information. This model sees visualization as a tool for externalizing thought processes and shaping the user's thinking in turn. Since we are dealing with a close interaction between cognition and visualization, the individual thinking style of a given user is likely to play a large part in the success and particular character of this process. In the end, understanding how visualization works for particular kinds of users will help us to understand the process of visualization itself.

References

[All00] ALLEN B.: Individual differences and the conundrums of user-centered design: Two experiments. *Journal of the American Society for Information Science 51*, 6 (2000), 508–520.

[Ama] Amazon Mechanical Turk, http://aws.amazon.com/mturk/. Retrieved 10-16-2008.

[Ber67] BERTIN J.: *Semiology of Graphics*. Univ. of Wisconsin Press, 1967.

[BH02] BREWER C., HARROWER M.: ColorBrewer, http://www.colorbrewer.org/, 2002. Retrieved 12-4-2008.

[Che00] CHEN C.: Individual differences in a spatial-semantic virtual environment. *Journal of the American Society for Information Science 51*, 6 (2000), 529–542.

[CM08] CONATI C., MACLAREN H.: Exploring the role of individual differences in information visualization. In *Advanced Visual Interfaces* (2008), ACM Press.

[DOBL06] DONNELLAN M. B., OSWALD F. L., BAIRD B. M., 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.

[EFHD76] ECKSTROM R. B., FRENCH J. W., HARMAN H. H., DERMAN D.: *Kit of factor-referenced cognitive tests*. Educational Testing Service, Princeton, NJ, 1976.

[Fek04] FEKETE J.-D.: The InfoVis Toolkit. In *Proceedings Information Visualization* (2004), IEEE CS Press, pp. 167–174.

[GG83] GENTNER D., GENTNER D. R.: *Mental Models*. Lawrence Erlbaum Associates, 1983, ch. Flowing Waters or Teeming Crowds: Mental Models of Electricity, pp. 99–129.

[GO90] GIBBS R. W., O'BRIEN J. E.: Idioms and mental imagery: the metaphorical motivation for idiomatic meaning. *Cognition 36*, 1 (1990), 35–68.

[Hal00] HALPERN D. F.: *Sex Differences in Cognitive Abilities*. Lawrence Erlbaum Associates, 2000.

[HPU98] HORN W., POPOW C., UNTERASINGER L.: Metaphor graphics to visualize ICU data over time. In *Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-98)* (1998).

[LJ80] LAKOFF G., JOHNSON M.: *Metaphors We Live By*. University of Chicago Press, 1980.

[MM99] MCQUARRIE E. F., MICK D. G.: Visual rhetoric in advertising: Text-interpretive, experimental, and reader-response analyses. *Journal of Consumer Research 26*, 1 (1999), 37–54.

[ZK08] ZIEMKIEWICZ C., KOSARA R.: The shaping of information by visual metaphors. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1269–1276.