Understanding the Interaction of Visual and Verbal Metaphors from Eye Gaze Behavior

Khairi Reda
University of Illinois at Chicago
mreda2@uic.edu

Dennis Chau
University of Illinois at Chicago
dchau1@uic.edu

Rohan Dargad
University of Illinois at Chicago
rdarga2@uic.edu

ABSTRACT

The visual metaphor framework proposed by Ziemkiewicz et al. characterizes visualizations by the visual metaphor they use to structure information. It postulates that viewers internalize the metaphor in their head and use it to conceptualize about the data depicted in the visualization. It also predicts that the way tasks are verbally worded affect the way viewers think about the visualization. In this paper, we aim to further characterize the interaction between visual and verbal metaphors by looking at eye gaze behavior of viewers. We study the visual attention of viewers when verbally primed with questions that are worded be either compatible with the visualization's visual metaphor or not. We found significant difference in the eye gaze behavior of subjects under the two different conditions. This difference suggests that participants had a better understanding of the visualization's structure when primed with a compatible task. This better understanding translated into more efficient and better-targeted fixations.

Keywords

Visualization theory, metaphors, hierarchies, cognition, eye-tracking.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Visual thinking, or the act of thinking with the aid of visual representation is a complex and multi-faceted phenomenon, involving coordination between multiple cognitive subsystems, including short-term visual and verbal memory, long-term conceptual memory, language processing centers, visual perception, and visual attention. The interplay of these different subsystems during visual thinking can not be ignored [Ware04, chapter 11]. Of particular interest is the interaction between visual and verbal/propositional subsystems, as most communication and decision making happens with the latter. Therefore, it is imperative for a science of visualization to provide a theoretical treatment of how these two cognitive subsystems.

Much of the theory underlying information visualization has focused on perception without much regard to high-level cognitive processes. The traditional practices of information visualization have been reduced to mapping a variable to a glyph [Bertin67]. This micro narrative has on one hand provided a good account of how users perceive individuals objects or data points within a visualization. However, this narrative fails to account for use cases in which users extract structural patterns and construct knowledge about the data depicted in the visualization. A good theoretical treatment will be beneficial to the practice of constructing novel and efficient visual representations for new types of data, as well as in evaluating existing designs. While user studies should remain an important step in visualization design, they do not in themselves necessarily constitute a sound scientific practice without a valid theoretical framework [Greenberg08].

One way to think about the nature of visualization is by considering the visual metaphor they use to structure information [Ziemkiewicz08]. For example, a node-link diagram depicting tree data uses levels as a metaphor for hierarchy. A treemap on the other hand uses containment as a metaphor for hierarchy. The two visualizations can be used to depict the same dataset. In this case, they are said to be

This paper has been presented as a course project in 2011 during a Human-Computer Interaction course at the University of Illinois at Chicago. The data was collected during course-related activities designed specifically for educational or teaching purposes as part of a class exercise.

"informationally equivalent". Yet it is obvious that there are major structural differences between the two. How do these differences affect the way people use and extract semantics from these two visualizations? Evaluations of node-link diagrams and treemaps have given conflicting results, with some finding treemaps better and others finding node-link diagrams better, even under very similar tasks. How can this be explained? Ziemkiewicz et al. propose that users internalize the visual metaphor in the visualization and use it to conceptualize the data. The internalized metaphor becomes the visualization's point of interaction with other cognitive subsystems such as language processing centers. This account suggests that the way a task is worded influence the performance of a user using the visualization. They suggest that tasks worded to be compatible with the visual metaphor embodied by the visualization illicit a more accurate response. For example, Consider the following two questions: "which of the two directories A, B, have more files directly under it" vs. "which of the two directories A, B contain more files directly in it". These two questions are equivalent: they are asking for the same piece of information. However, the former is worded with a "levels" metaphor, making it compatible with node-link diagrams, where as the latter is using a "containment" metaphor which is more compatible with treemaps. Ziemkiewicz et al. experimented with the four combinations (node-link, treemap) x (levels metaphor, containment metaphor) [Ziemkiewicz09]. They found that questions worded in a verbal metaphor that is compatible with the visualization's yielded higher overall accuracy. Ziemkiewicz et al's experiment indeed suggests an interaction between verbal and visual metaphors. However, the experiment also raises questions about the nature of that interaction. Can this interaction be characterized more precisely and attribute to one or more cognitive processes? One way to investigate this interaction is by looking at the visual attention of participants, and how that allocation differs across the two metaphor compatibility conditions.

In this paper, we aim to elaborate on the interaction of visual and verbal metaphors by studying how viewers allocate their visual attention when answering questions about data depicted in the visualization. We report on a user study similar Ziemkiewics et al's experiment with the addition of an eye tracker to study the gaze behavior of participants. Our results show that there is a significant difference in the way participants allocate their visual attention to think visually when presented with a verbal statement that is either compatible or incompatible with the visualization's inherent visual metaphor. Our analysis of gaze behavior suggests that viewers have a better understanding of the visualization structure when primed with a compatible metaphor, and can move their eyes rapidly between the relevant sections of the visualization and make better targeted fixations.

RELATED WORK

Diagrammatic reasoning has been an active area of research in psychology circles and has had a long history. Larkin and Simon outline the essential differences between sentential (text) and visual representations [Larkin87]. They assert that visual representation make salient elements that are essential for problem solving, but yet are implicit in sentential description. Thus, diagrams and visualization allow for a different set (presumably more powerful or diverse) of inference rules to be applied to solve the underlying problem.

Another influential work comes from Pinker [Pinker90]. Pinker postulates a graph schema, an implicit knowledge that a reader uses to read a graph. Pinker's model can be used to explain how people extract a specific piece of data from a graph to answer a well-defined quantitative question about the data. However, it cannot be used to explain how people interpret the data as a whole to provide a qualitative treatment of the information embodied by the graph. Zacks and Tversky show that the overall structure of a graph provide the viewer with a framework for interpreting the data [Zacks99]. For example, bar charts suggest a discrete set of categories whereas line charts suggest to the viewer the data points are related and should be assessed as a trend. A viewer's assessment of a graph was remarkably consistent with this account even when line charts where used to depict clearly discrete categories (such as gender), in which case the viewers responded with odd assessment such as “the more Male a person is, the taller he is”. This account can be regarded as an extension to Pinker’s graph reader’s schema, with additional interpretation frameworks associated with the overall structure of the graph added to Pinker’s schema. It is not clear though how this account can be generalized to InfoVis representations which typically employ novel visual encoding and structure.

Zacks and Tversky’s account presumes the interpretation framework is reinforced (and perhaps established) by communication standards (line charts being used to illustrate financial trends in newspapers, for examples). However, novel InfoVis representations are unlikely to have such a strong interpretation framework. In this case, the viewer could rely on a more general interpretation framework. The notion of a visual metaphor could be used to provide such a framework. Ziemkiewicz et al’s suggest that the visual metaphor can be considered to be the essence of the visualization [Ziemkiewicz08]. This visual metaphor is simulated and applied to the information embedded in the visualization so they can be conceptualized. This account can be seen as a parallel to Lakoff et al’s view on languages (and the human mind, for that matter) as being deeply metaphorical in their nature [Lakoff80].

METHODS

We designed a similar experiment to Ziemkiewicz et al’s [Ziemkiewicz09] in order to shed further light on the interaction between visual and verbal metaphors. Specifically, our goal is to further characterize that interaction at the visual attention level, and explain the underlying reason for it. We use two visualizations of hierarchical data, node-link diagrams and treemaps, and expose the participants to problems that were verbally framed to be either compatible or incompatible with the visualization’s metaphor. A node-link diagram is assumed to use a hierarchy-by-levels metaphor, whereas a treemap is assumed to use a hierarchy-by-containment metaphor. Unlike Ziemkiewicz’ experiment; we do not evaluate the spatial ability of participants. However, in addition to analyzing answers and response times, we also track the participants’ eyes during the experiment and analyze their eye gaze behavior.

Participants

We recruited 13 participants to undertake our experiment. Participants were seated in front of a 17-inch LCD monitor. Two infrared cameras were attached underneath the monitor to track the pupil of the participant during the experiment. The location of the eye fixations are recorded to a log file for later analysis. Two of the participants were dropped from the eye gaze behavior analysis due to their excessive head movement. However, their accuracy and response time results were included in the analysis.

Procedures

We prepared 24 stimuli broken into two blocks of 12 stimuli each. Each stimulus is composed of a question followed by the visualization. One block presents questions that are worded to be compatible with the visualization’s metaphor; where as the other block presents incompatible questions. Within each block, there are six node-links diagrams and six treemaps, making up the 12 stimuli in the block. The two blocks contain identical stimuli. However, the wording of the questions varies to reflect either a compatible or an incompatible metaphor. Additionally, the names of the nodes in the visualizations were changed, and the tree structures were shuffled to conceal the similarity between the two blocks and eliminate learning. This design uses a within-subject treatment; each subject is exposed to two equivalent blocks (12 stimuli each) with one block of tasks presented using a verbally compatible wording, whereas the second block presents tasks verbally to be incompatible with the visualization’s metaphor.

The experiment comprises a series of 24 stimuli in a random order. Each stimuli consists first of a question (task) followed by a visualization. The subject first seems the question. When he/she is finished reading the question, the subject presses the space bar at which point the question disappears and the visualization appears. The subject answers the question by pressing ‘Q’ for Yes or ‘P’ for No. After this, the visualization disappears and a circle appears in the center of the screen for three seconds before the next stimuli is displayed. The experiment program logs the time the subject took to looking at each stimuli, as well as the responses of the subject.

Before the 24 stimuli are presented, a series of 6 stimuli are presented first for training. In the training phase, the subject can move to the next stimulus only after answering the question correctly. An incorrect answer results in a quick red flicker of the screen to indicate an incorrect response, in which case the subject will try again.

Tasks

The tasks are presented in the form of a question which is displayed before the visualization. We used three types of questions identical to the ones used in Ziemkiewicz et al's experiment [Ziemkiewicz09]. Table 1 lists examples of these questions worded in a 'hierarchy-by-containment' as well as a 'hierarchy-by-levels' metaphor.

Containment metaphor	Levels metaphor
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?

Table 1. Sample questions with a containment metaphor on the left and a levels metaphor on the right.

RESULTS

Results were analyzed to determine the effect of metaphor compatibility on accuracy, response time, and eye gaze behavior.

Accuracy

We collapsed the accuracy data into 2 categorical variables (compatible vs. incompatible metaphors) and counted the number of correct and incorrect responses under each category. We measured the effect of metaphor compatibility on the total number of correct and incorrect responses. Based on Ziemkiewicz et al's results, we hypothesized that the compatible condition will yield a higher number of correct responses.

Figure 1 shows the total number of correct and incorrect responses under the two conditions. Contrary to our hypothesis, we see a higher number of correct responses when tasks were worded in an incompatible metaphor. However, a chi-square test did not find any significant difference between the two treatment groups, $\chi^2(1, N = 312) = 1.701, p > 0.05$. One possible explanation for this result is that our tasks were easier than the tasks in Ziemkiewicz et al's study. This was due to the fact that we limited our tree datasets to a maximum depth of 6 levels, whereas their experiment had datasets with a maximum depth of 8 levels.

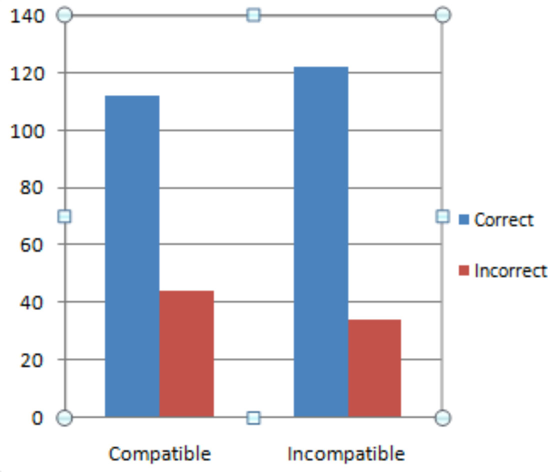

Figure 1. Total number correct vs. incorrect responses for the compatible and incompatible conditions

Response time

The average response time a subject spends looking at the visualization portion of the stimulus was calculated under the two compatibility conditions. The response times per stimulus ranged from 2.62 to 32.74 seconds with mean response time for a stimuli being 10.2 (S.D. = 5.54). Figure 2 shows the average time a participant takes to solve a compatible vs. an incompatible block. This is consistent with Ziemkiewicz et al's results. However, their results were not statistically significant. In our case, a two-tailed t-test (P=0.0932) (t value = 1.82) confirms that participants spend significantly more time looking at a visualization when verbally primed with an incompatible metaphor than with a compatible one.

The longer response time indicates a higher cognitive workload in the incompatible block. Since the two blocks are identical in their difficulty, this suggests that the additional cognitive workload is due to the need to translate between the verbal and visual metaphors in the incompatible block.

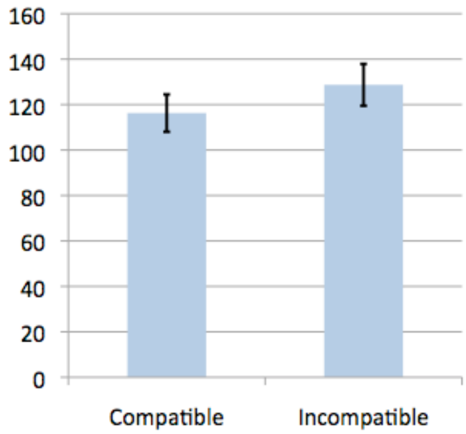

Figure 2. Average time (in seconds) to complete the compatible vs. the incompatible block (12 stimuli each). In all figures, error bars represent the standard error.

Eye gaze behavior

Average number fixations

An eye fixation is perhaps the most commonly used metric in eye tracking analysis. A fixation refers to locations on the screen where the participants focused their visual attention on. A fixation was defined as a minimum of five eye samples (~ 100 ms) within 12 pixels (approx 2° of visual angle) [Ratwani08]. Average number of fixations per stimulus was calculated for each participant and the total was then averaged separately for the two compatibility conditions. Figure 3 shows that average number of fixations for a compatible vs. incompatible stimulus. A two-tailed t-test (P=0.076) (t value = 2.02) confirms that incompatible metaphors illicit a larger number of fixations from participants. A higher number of fixations confirm that participants took more time when interpreting the visualization after being primed with an incompatible metaphor.

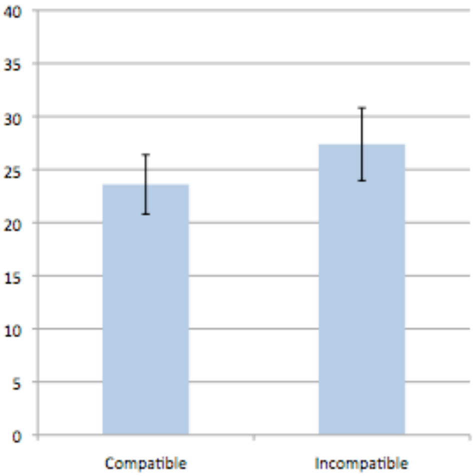

Figure 3. Average number of fixations per stimulus.

Average fixation time

Longer fixations indicate that the participant is having difficulty extracting information from the visualization [Fitts50, Goldberg98]. That is, difficulty in perceiving the individual components of the visualization (for example, nodes in a node-link diagram). Figure 4 show the average fixation time under the two compatibility conditions. We did not find any significant difference between the two groups. This indicates that the individual components of the visualization (nodes) are equally perceived under both conditions, suggesting that the additional overhead in incompatible stimuli was due to higher-level cognitive process, as opposed to difficulty in perceiving the elements of the visualization. Moreover, a look at the time of an average fixation per stimuli revealed that 99% of them fell below the 240ms mark. That is, they were involuntary fixations [Graf89]. This tells us that with respect to our study, it was relatively easy to extract information from the visualization irrespective of the compatibility condition.

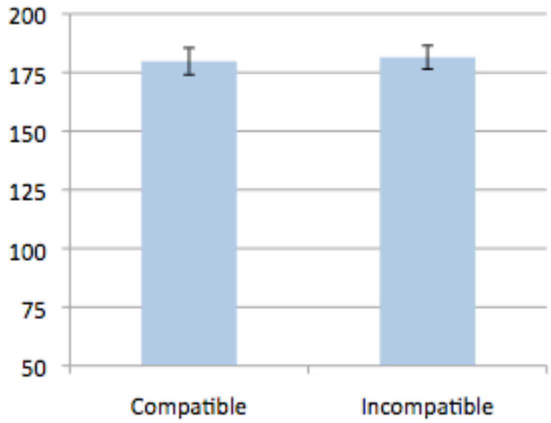

Figure 4. Average fixation time (in Milli seconds).

Saccadic amplitude

Saccadic amplitude is the distance in pixels (or visual arc angles) between two successive fixations [Carpenter88]. We hypothesized that higher saccadic amplitudes correlate with confusion as they would indicate jumping erratically in a non-systematic way over the visualization. Thus we expected lower saccadic amplitudes in compatible questions.

We calculated the average saccadic amplitude for each stimulus, which was averaged separately for each of the two compatibility conditions (Figure 5). Results showed significant difference between the two conditions as confirmed by a two-tailed t-test (P = 0.0018) (t-value = 4.2). However, contrary to our hypothesis, the saccadic amplitude was higher for the compatible condition. We believe this was due to the fact that, in the compatible condition, participants had a better understanding of the visualization's structure and could anticipate were the required information was spatially located within the visualization. Thus participants were able to make efficient, well-targeted jumps in compatible stimuli, resulting in higher average saccadic amplitudes.

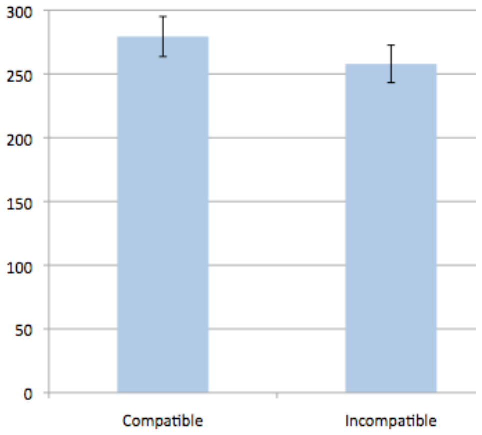

Figure 5. Average saccadic amplitudes (in pixels).

Coded analysis

The goal of coded analysis is to associate fixations with high-level features of the visualization. The frequency of the high level features are then analyzed to get insight into why a participant makes certain associations when attempting to answer the questions under a given metaphor condition. We based our coding scheme on the work of Ratwani et al. [Ratwani08]. Since the tasks in our experiment were questions about relationships between nodes, we designed a coding scheme to capture the participant's effort to form relations between the nodes in the visualization. This includes child-parent relations and sibling relations (two or more nodes sharing the same parent). The fixations were either coded to be relationship forming, or non-relationship forming. The percentage of relationship forming fixations from the total number of fixations was calculated and averaged separately for the two compatibility conditions.

The criteria for coding a fixation began with the location of the fixation: on a node, on a link/border, in an unknown location, etc. Afterward the previous location was compared with the current in order to determine if a relationship between the two fixations could be formed. This coding scheme required the manual playback and coding of the fixations, which took approximately one hour for each participant. Because of the difference between node-link diagrams and treemaps, we developed a separate, though similar coding scheme for each. Figure 6 and 7 illustrate the coding scheme for node-link diagrams and treemaps, respectively.

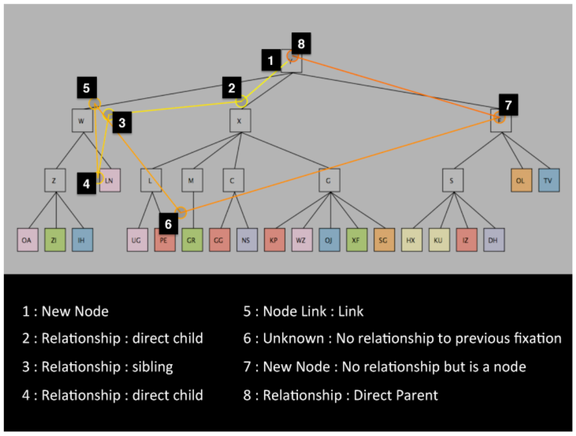

Figure 6. An example illustrating the coding for a node-link diagram. The circles indicate fixations and the line indicate saccade trajectories. Lines were colored using in a gradient to illustrate ordering from yellow (earlier) to red (further in time). At the bottom of the figure is the assigned code for each fixation.

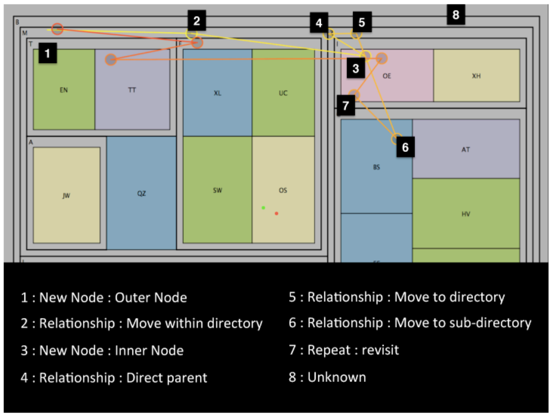

Figure 7. An example illustrating the coding for a treemap.

We expected that an incompatible metaphor will cause more confusion in interpreting the visualization. This confusion will manifest itself with a more "scattered" fixation patterns that are less systematic. Thus, we expected a significantly lower percentage of relationship-forming fixations in incompatible. Figure 8 shows the percentage of relation-forming fixations for node-link diagrams and treemaps under either a compatible or incompatible metaphor. Contrary to our hypothesis, a t-test did not find significant difference in the percentage of relationship-forming fixations under the two metaphor conditions.

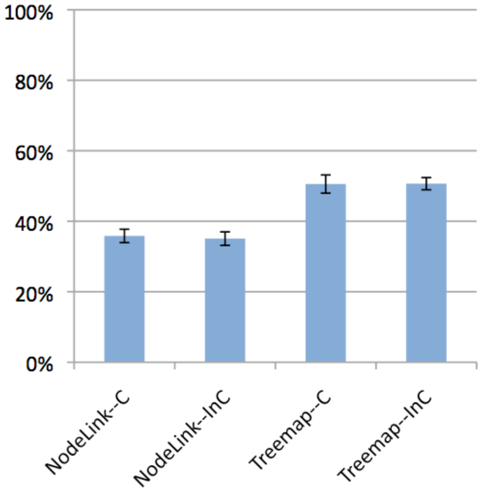

Figure 8. Percentage of Relational Transition for node-link diagrams and treemaps, under either a compatible (C), or incompatible (InC) metaphor.

DISCUSSION

Our analysis of the accuracy of responses did not reveal significant interaction between metaphor compatibility and the probability of answering the question correctly. While this is somewhat surprising and contradictory to Ziemkiewicz et al's results, this could be due to two reasons. First, our sample size (13 participants) was significantly smaller than the sample size of Ziemkiewicz et al's (63 participants). The second factor could be related to the difficulty-distribution our questions. After analyzing the responses of participants we found that about two thirds of the questions were answered correctly by most participants, suggesting that these questions were easy. Two other questions resulted in zero correct responses, indicating that they were very difficult to answer under both compatibility conditions (possibly due to the tree dataset being particularly confusing). We expect difficult questions to be more susceptible to adverse interaction from incompatible metaphors. However, we suspect that most questions used in our experiment were either too easy or too difficult to answer, diminishing the interaction between incompatibility and accuracy.

Contrary to Ziemkiewicz et al., our results show a significant interaction between compatibility and response time (the time a participant spends looking at the visualization before answering the question). Participants spent roughly 10% more time looking at the visualization when the questions were framed in an incompatible metaphor. While the difference is not huge, it is significant. This difference suggests a higher cognitive work load which could be due to participant's effort in translating between the verbal and visual metaphors. The results of eye gaze behavior also support this hypothesis. We see a larger number of fixations in incompatible questions, while the time length of a single fixation remains constant under both compatibility conditions. This constancy in fixation time suggests that perceptual workload in extracting individual information elements form the visualization seems to be constant across the two conditions, as fixation time is correlated with difficulty in extracting visual information. This indicates that the additional cognitive workload due to metaphor incompatibility is more likely to be at the conceptual than at the perceptual level. Additional support for this hypothesis comes from the saccadic amplitudes. We see significantly larger saccadic amplitudes in compatible stimuli (Figure 5). This, in conjunction with the lower response time suggest that participants have a better understanding of the visual structure, and can anticipate more accurately where to find the relevant pieces of data that are required to answer the question. Thus, when primed with a compatible metaphor, a participant can make more efficient saccades and well-targeted fixations to acquire the necessary pieces of data.

It is also interesting that Ziemkiewicz et al. did not get an increase in response time under incompatible metaphors. This could be explained by the way our stimuli set was organized. As explained in section 3.2, we used two blocks that had either compatible or incompatible questions containing 12 stimuli each. Additionally, the tree datasets, the questions, and the nodes that the questions referred to were identical in both blocks. Only the names of the nodes and the structure of the tree were shuffled to conceal similarity. This resulted in two blocks that were equal with respect to stimuli difficulty. The within subject design of the experiment and the equal difficulty perhaps facilitated the emergence of small but significant differences in timing. Moreover, we randomized the order of stimuli presentation for each participant, which could have probably averaged any learning effect, helping the differences in response time emerge at the block level.

The coded analysis did not reveal any significant difference between compatible and incompatible stimuli. However one interesting finding was a difference in the percentage of relation-forming fixations between node-link diagrams and treemaps. The additional 20% relation-forming fixations in treemaps indicate that interpreting relations was harder in treemaps. This was also confirmed from qualitative comments by participants. We see this as a positive indication that the coding scheme is capturing some part of the cognitive process. However, the current coding scheme failed to capture the essential differences due to metaphor incompatibility. One way to capture this in a future coding scheme is to encode larger and higher level fixation-saccade patterns. Figure 9 and 10 compares saccade trajectories under an incompatible and a compatible metaphor, respectively, for the same participant. Figure 9 shows “Looping Patterns” in blue, where as Figure 10 shows movements that appear to be more horizontal in nature. Taking these high-level differences into account in the coding future may help reveal further differences in the participant's cognitive effort under the two metaphor compatibility conditions.

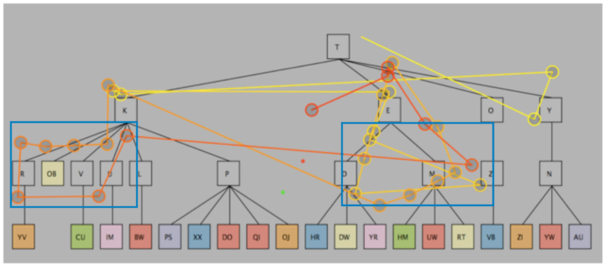

Figure 9. An example of a participant's fixations and saccades under an incompatible metaphor. This figure illustrates the "Looping Patterns" (blue rectangles)

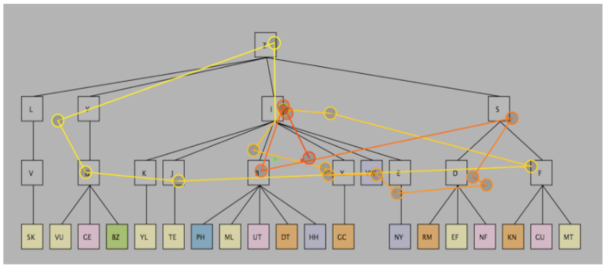

Figure 10. An example of a participant's fixations and saccades under a compatible metaphor illustrating the horizontal saccade pattern.

CONCLUSIONS

The visual metaphor framework characterizes visualizations by the visual metaphor they use to structure information. It postulates that viewers looking at a visualization internalize the metaphor in their head and use it to conceptualize about the data depicted in the visualization. This internalized representation also interacts with language processing centers. This interaction results in lower response accuracy when tasks are worded in a verbally incompatible metaphor with the visualization's visual metaphor. In this paper, we shed a further light on this phenomenon and provided characterization of the interaction between the visual and verbal metaphor at the visual attention level. Analysis of gaze behavior suggests that viewers have a better understanding of the visualization structure when primed with a compatible metaphor, and can move their eyes rapidly between the relevant sections of the visualization and make better targeted fixations. This extra efficiency for compatible metaphors results in lower response time and potentially higher response accuracy.

REFERENCES

[Bertin67] Jacques Bertin. Semiology of Graphics. Gauthier-Villars, 1967

[Carpenter88] R.Carpenter. Movements of the Eyes. London: Pion, 1988

[Fitts50] P. Fitts, R. Jones, and J. Milton. Eye Movement of Aircraft Pilots during Instrument-Landing Approaches. Aeronautical Engineering Review 9, 24-29, 1950

[Goldberg98] J. Goldberg and X. Kotval. Eye Movement-Based Evaluation of the Computer Interface. In Kumar, S. K. (Ed.): Advances in Occupational Ergonomics and Safety, Amsterdam: IOS Press, pp. 529-532, 1998.

[Graf89] W. Graf and H. Krueger. Ergonomic evaluation of user-interfaces by means of eye-movement data. In Proceedings of the third international conference on human-computer interaction, 1989

[Greenberg08] S. Greenberg and B. Buxton. Usability evaluation considered harmful (sometimes). In Proc. of the SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, 2008.

[Lakoff80] G. Lakoff and M. Johnson. The metaphorical structure of the human conceptual system. Cognitive Science, 4:195–208, 1980

[Larkin87] J. Larkin and H. Simon. Why a diagram is (sometimes) worth ten thousand words. Cognitive Science 11, pp. 65-99, 1987.S

[Pinker90] Steven Pinker. A theory of graph comprehension. Artificial Intelligence and Future of Testing. 1990.

[Poole05] A. Poole and L. Ball. Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. Chapter in C. Ghaoui (Ed.): Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group, Inc, 2005

[Ratwani08] R. Ratwani, J. Trafton, and D. Boehm-Davis. Thinking Graphically: Connecting Vision and Cognition During Graph Comprehension. Journal of Experimental Psychology: Applied. 14(1):36-49, 2008

[Shah09] P. Shah and E. Freedman, Bar and Line Graph Comprehension: An Interaction of Top-Down and Bottom-Up Processes. Topics in Cognitive Science, 2009

[Ware04] Collin Ware. Information Visualization: Perception for Design, second edition. Morgan Kaufmann, 2004

[Zacks99] J. Zacks and B. Tversky. Bars and Lines: A study of graphic communication. Memory and Cognition 27, pp. 1073-1079, 1999

[Ziemkiewicz08] C. Ziemkiewicz and R. Kosara. The shaping of information by visual metaphor. IEEE Transactions on Visualization and Computer Graphics, 2008

[Ziemkiewicz09] C. Ziemkiewicz and R. Kosara. Preconceptions and Individuals Differences in Understanding Visual Metaphor. In Proc. Eurographics Symposium on Visualizaiton, 2009