AN EMPIRICAL STUDY OF THE EFFECTS OF GESTALT PRINCIPLES ON DIAGRAM UNDERSTANDABILITY

Krystle Lemon[1], Edward B. Allen[1], Jeffrey C. Carver[1], and Gary L.Bradshaw[2]

[1]Department of Computer Science and Engineering
300 Butler Hall, Box 9637
Mississippi State, MS 39762
+1 662-325-2756
{kdl18, carver}@cse.msstate.edu;
edward.allen@computer.org

[2]Department of Psychology
Box 6161
Mississippi State, MS 39762
+1 662-325-0550
glb2@ra.msstate.edu

Abstract

Comprehension errors in software design must be detected at their origin to avoid propagation into later portions of the software lifecycle and also the final system. This research synthesizes software engineering and Gestalt principles of similarity, proximity, continuity for the purpose of discovering whether certain visual attributes of diagrams can affect the accuracy and efficiency of understanding the diagram. The experiment tested whether two dependent variables, accuracy and response time, were significantly affected by independent variables, diagram type (simple1, simple2, complex), Gestalt principles (good vs. bad), and question order (forward/backward). The results of this study indicated that the Gestalt principles did affect the comprehension in the complex diagrams. Post-hoc analysis results indicated that number of bends per line, length of line in inches, number of lines crossing, boxes per diagram, and number of lines per diagram contributed to the ability of the subjects to comprehend the diagrams.

Keywords

Gestalt principles, diagram comprehension, empirical software engineering, software architecture, cognitive science

1. Introduction

Software engineers have the tasks of designing, coding, and testing the software systems that they build. Software architecture diagrams, like blueprints, depict designs of software and guide subsequent phases. These diagrams depict connections and components of the system and describe the messages and behavior of the system to be implemented. Our goal is to understand errors in interpretation of software architecture diagrams in order to help software engineers avoid mistakes and pitfalls. Understanding and correctly implementing the software architecture is a key factor in building the correct system and fulfilling the customer's requirements. It is common wisdom in software engineering that it is more costly to correct errors later in the software lifecycle than in the earlier phases. Avoiding errors can indeed save valuable time and money on software projects.

The ultimate goal of this research is to understand what errors are made in software diagram comprehension and what the roots of these errors are. Cognitive science is a discipline that studies the human mind and how it works [6]. This field of study coupled with software engineering can help to gain more insight into comprehension of software engineering diagrams. Building on work reported in the literature, this research investigates attributes of diagrams that may affect comprehension. In cognitive science, Gestalt principles of perceptual organization deal with features that combine to form overall perception, such as relationships among visual features. Details are provided in Section 2. Research that explicitly applies Gestalt principles has been used in the psychology field, but these principles have not been applied to software-engineering UML diagrams. This paper presents the results of a study coupling cognitive science and software engineering to investigate whether certain diagram characteristics affect comprehension by software engineers. Some preliminary results were reported in an earlier short paper [5].

This paper is organized as follows. Section 2 provides an overview of diagram comprehension and related work. Section 3 describes the study. Section 4 contains the results and discussion and post-hoc analysis. Lastly, Section 5 discusses the conclusions and future work.

2. Background and Related Work

This section presents some previous work on diagram comprehension and their results. It describes some of the research done to identify factors that influence diagrammatic comprehension.

2.1 Diagram Comprehension

Software-architecture diagrams show the flow of information between components in a software system. Such diagrams consist of lines with arrowheads pointing in the direction of information flow, boxes representing system components, and other annotations. Software engineers must comprehend all of this information and process it within their mental workspace [4]. Hungerford, showed that diagrams impose less cognitive load compared to text [3]. Not surprisingly, software engineers routinely make extensive use of a variety of diagrams.

Pioneering research conducted by Purchase et al. [7] investigated graph layout algorithms and their effect on diagram comprehension. Their work was domain-independent, and in particular, it did not apply to software engineering or any other specific engineering disciplines. They used sparsely and densely populated diagrams that contained unlabeled nodes and edges. After the initial work, Purchase et al. [9] performed additional empirical studies of how the drawing aesthetics and syntax of UML diagrams affect users' comprehension. This work studied graph aesthetics of edge bends, edge crosses, maximizing the minimum angle of edges, orthogonality and symmetry between different pairs of graphs and how subjects performed on certain tasks that measured their degree of comprehension. Results from this work showed that the attribute of crossing lines was the most important aesthetic to control, because users' comprehension of the UML diagrams decreased when the number of crossing lines increased. Other aesthetic attributes were measured but did not significantly affect the comprehension of the diagrams by the subjects.

Additional studies have emphasized other graph aesthetics not heavily studied in prior research. Ware et al. [11] explored the problem of path continuity and its effect on diagrammatic reasoning and comprehension. The findings of this study suggest that the principle of continuity aids in recognizing the shortest path in a graph. This recognition could lead to a faster response time in comprehending the diagram.

2.2 Gestalt Principles

The Gestalt principles, which originated in the field of psychology, address why individuals can perceive whole elements out of incomplete elements. Some of the issues they address are how objects are viewed in relation to similarity, proximity, continuity, closure, area, and symmetry [1]. Gestalt principles have been used by psychologists on many occasions to demonstrate how humans view or perceive the world based on the organization of objects. Psychologists have gone so far as to say that without Gestalt principles of organization humans would view the world chaotically and without organization [2].

The study described in this paper focuses on the Gestalt principles of similarity, proximity, and continuity. The principle of *proximity* states that objects that are physically close together are perceived as belonging to the same group or set. In Figure 1, the picture displays the Gestalt principle of proximity that suggests the picture is perceived as 2 sets of columns of squares as opposed to 4 columns of squares. The principle of *similarity* states that objects with similar characteristics are perceived as belonging together. In Figure 2, the picture represents the Gestalt principle of similarity that suggests that the squares are viewed as 3 rows of small squares and 1 row of larger squares instead of 4 columns of squares. Finally, the principle of *continuity* states that continuous figures are comprehended more easily and more quickly than non-continuous figures. An example of a continuous figure is when a line is perceived to pass through an object instead of viewing it as two separate lines on the object: one entering and another leaving. Figure 3 suggests that the figure is perceived as two lines crossing as opposed to 4 lines meeting at the center.

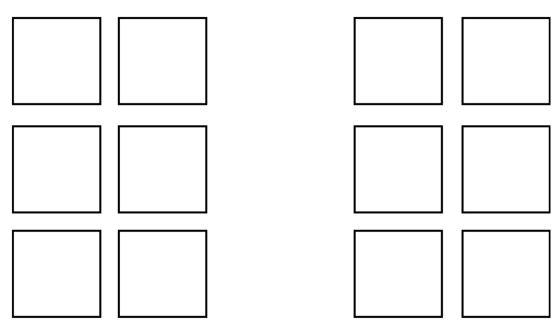

Figure 1. Proximity

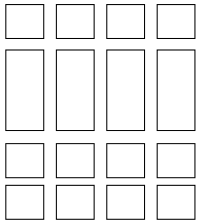

Figure 2. Similarity of Size

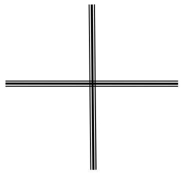

Figure 3. Continuity

Dodd and Pratt [2], examined how attention was allocated to objects grouped by Gestalt principles. The experiment tested whether grouped versus non-grouped objects had faster and more accurate response times when cued by particular stimuli. This study helped to strengthen the argument that humans naturally give attention to objects grouped by Gestalt principles.

3. The Study

Section 3.1 details the research hypotheses and Sections 3.2 and 3.3 describe the subjects and variables, respectively. Section 3.4 describes the data collection efforts.

3.1 Hypotheses

This initial study did not try to separate out the effects of the different Gestalt principles; rather they were all taken together. This study focused on two hypotheses, in the context of diagram comprehension:

Hypothesis 1: Diagrams that follow Gestalt principles of similarity, proximity, and continuity offer better accuracy than diagrams that do not.

Hypothesis 2: Diagrams that follow Gestalt principles of similarity, proximity, and continuity offer faster response time than diagrams that do not.

3.2 Variables

The study had two dependent variables and three independent variables. The dependent variables investigated in this study were *question accuracy* and *response time*. The independent variables were: *good/bad* (Gestalt principles used), *type of diagram* (*simple1*, *simple2*, or *complex*), and *question order* (forward/backward). The question order was used as a control variable to determine whether the presentation order of questions had any effect on subject response.

3.3 Study Design

The study was conducted at Mississippi State University in the Fall of 2005. The subjects were randomly split into four groups with approximately the same number of subjects as shown in Table 1. Further explanation is provided in Section 4.2. This table indicates what type of diagram was seen by the subjects and the ordering of the questions.

Table 1: Experiment Design

Group	Diagram Order	Question Order	No. of Subjects
1	Good/Bad/Good	Forward	6
2	Good/Bad/Good	Backward	6
3	Bad/Good/Bad	Forward	7
4	Bad/Good/Bad	Backward	7

3.3.1 Subjects. The 27 subjects were drawn from a Software Architecture course (15 subjects) and an Introduction to Software Engineering course (12 subjects). These subjects were upper-level undergraduates and graduate students. The software architecture course covers basic software architecture concepts including the creation and use of architecture diagrams. The Introduction to Software Engineering course focuses on the software engineering process and the use of diagrams during that process. The subjects' participation in the study was part of their final grade but they did not receive a grade for how well they performed.

3.3.2 Artifacts. To simulate real-world diagrams, three diagrams, created by software-engineering students as part of a homework exercise during a previous semester of the Software Architecture course, were selected for use in the study. Because the homework diagrams by the class were similar, three diagrams were selected for use in the study. Each diagram contained named boxes (system components) and lines with arrowheads (representing information flow). The three diagrams were labeled *simple1*, *simple2*, and *complex*, generally indicating the number of boxes on the diagram and the number of lines between boxes. For each of the three diagrams, two topologically equivalent versions were used. The bad version of each diagram was simply the original diagram. The

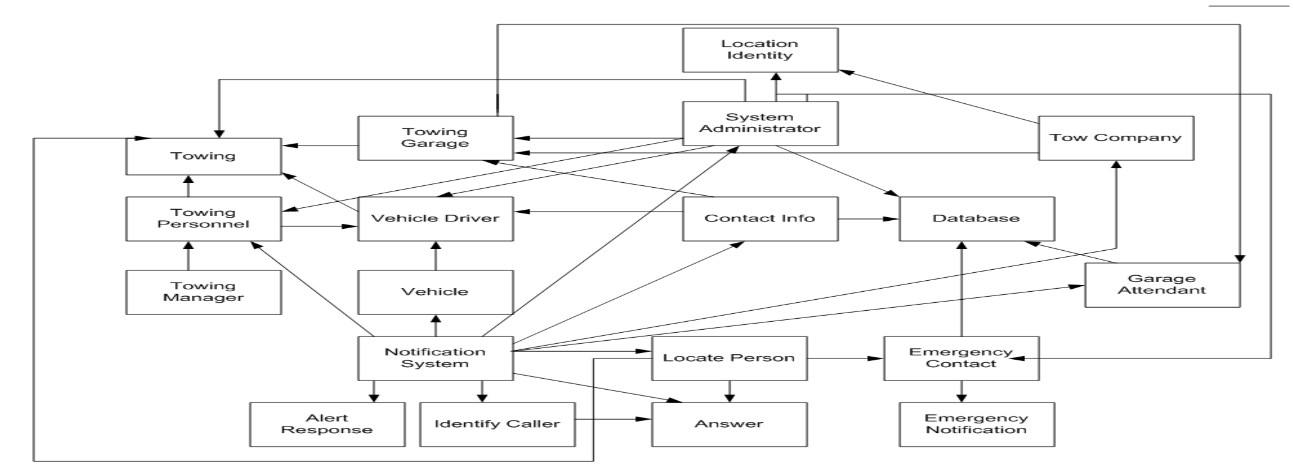

Figure 4. Complex diagram good version

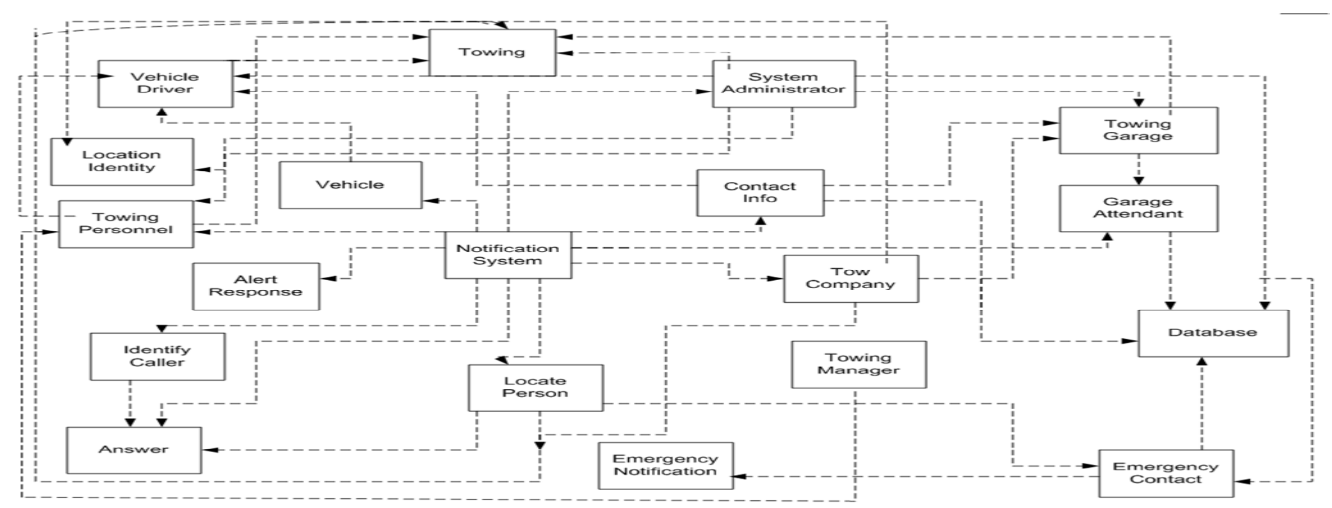

Figure 5. Complex diagram bad version

good version of the diagram was created by modifying the bad version using the Gestalt principles (described in Section 2). The modifications included: dashed lines were changed to solid lines, boxes that had similar names were grouped together, and spacing between boxes was minimized. Figure 4 and 5 show the good and bad versions of the complex diagram.

3.3.3 Experimental Procedure. All subjects viewed the diagrams in the same order, *simple1* diagram followed by a *simple2* diagram followed by a *complex* diagram. For each diagram, half of the subjects viewed the *good* version while the other half viewed the *bad* version. Each subject was asked 20 questions per diagram (the same questions were used for both good and bad diagrams). The subjects were given as much time as needed to answer the questions. At their own pace, they proceeded to answer the questions one by one while the associated diagram was still visible. They were not allowed to skip any questions. The questions asked the subjects to determine whether messages were sent between objects in the diagram (i.e. whether there was a line connecting the boxes). An example question is “Does Location Identity send messages to Towing Garage?” To assess the potential affects, some subjects answered the questions in the *forward* order (i.e. 1-20) while some answered the questions in *backward* order (i.e. 20-1).

3.4 Data Collection

The Macromedia Authorware software package was used to collect the results of the study which included the *accuracy* and *response time* variables.

4. Results and Discussion

This section describes the statistical analysis performed on the quantitative data in order to determine whether a significant difference in diagram comprehension existed between different types of diagrams. As this was an initial study, an alpha value of 0.1 was chosen for significance tests.

4.1 Summary of Quantitative Data

The main goal of the statistical analysis was to determine if the Gestalt principles had a significant effect on either of the dependent variables (accuracy or time). The independent variables (diagram type and question order) were analyzed to understand their effects on the results.

Prior to performing the main analysis, an outlier analysis was conducted. One subject's accuracy was more than two standard deviations below the mean, so the data from that subject was excluded from all further analysis, leaving data from 26 subjects.

During the analysis of the data, all of the questions were renumbered based on the diagrams they referred to. The questions about *simple1* were numbered 101 through 120, the questions about *simple2* were numbered 201 through 220, and the questions about complex were numbered 301 through 320. It was discovered that different versions of a few questions were used for the good diagram than were used for the bad diagram. To prevent any bias, two questions were discarded. The questions that were discarded are not present in the graphs and were not used in the analysis. (Due to this fact, some of the diagrams are missing question ID numbers).

For the dependent variable *question accuracy,* most subjects were very accurate with low variability cross all diagrams. The mean was 18.65 correct answers per diagram with a standard deviation of 1.22 answers. The lack of variation in this variable indicated that further analysis of this variable was unlikely to show interesting results. Therefore, the remainder of the analysis focuses on the *response time* variable.

A 3-Way ANOVA test was conducted separately for each of the three diagrams (i.e. *simple1*, *simple2*, and *complex*). This analysis was performed to isolate the effects of the three independent variables for each diagram. The ANOVA test for response time for *simple1* diagram did not show a significant overall effect. However, it did indicate a significant interaction between the *question order* and *good/bad* variables with $F_{(27,1)}$ =5.896 and *p*=0.023. A further analysis was done using independent sets *t*-test to isolate the effects of the *forward/backward* variable and the *good/bad* variable. Neither of these variables individually had a significant impact on the response time.

The ANOVA test for response time in the *simple2* diagram yielded no significant results for any of the variables. Finally, the ANOVA test for response time on the complex diagram showed that the *good/bad variable* had a significant impact on response time ($F_{(27, 1)}$ = 0.001 and *p*=0.004).

4.2 Discussion of Quantitative Results

The high percentage of questions answered correctly by all subjects suggests that the diagrams and/or questions were not complex enough in the study to adequately evaluate Hypothesis 1 related to accuracy.

Figure 6 summarizes the distribution of response times across the three diagrams. The analysis shows that while the Gestalt principles did not have an effect on the subjects' response time for the simple diagrams, they did have a significant impact on the complex one. Therefore, this result provides some support for Hypothesis 2 that comprehension time can be reduced if Gestalt principles are used. This conclusion motivates additional study to better understand which Gestalt principles can be applied in software engineering to aid in diagram comprehension and which should be avoided.

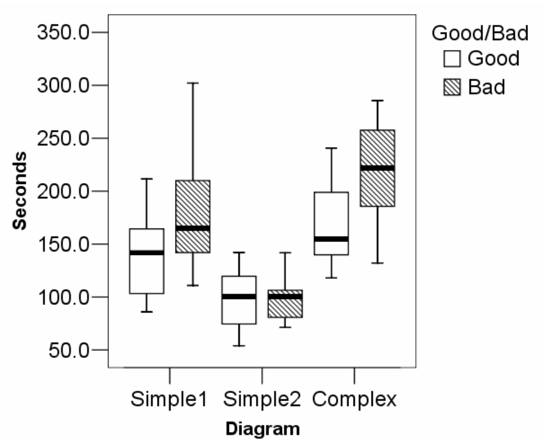

Figure 6. Total response time

4.3 Post Hoc Analysis

Strong effects between different attributes of the diagrams were noticed in the initial results. The post-hoc analysis was designed to probe into those effects to discover what influenced their presence. The response time variable was broken down into response time per question for additional analysis. The response time for all subjects was averaged for each question in combination with the forward or backward direction of the questions. This analysis resulted in two average times for each of the questions (one for the subjects who viewed the questions in the forward order and one for the subjects who viewed the questions in the backward order). Visual and quantitative comparisons could then be made for each diagram and question.

The goal of this analysis was to understand if certain types of questions caused slower reaction times (i.e. questions that contained boxes far from each other, or questions that had boxes that were not grouped with similar boxes). Figure 7 is a line graph that represents the average reaction time of each question in the *simple1 good* diagram *forward* direction. As more questions were seen, response times became faster. The same downward trend is present for the subjects who answered the questions in backward order. This result indicated that question order did not influence response time more than any particular question.

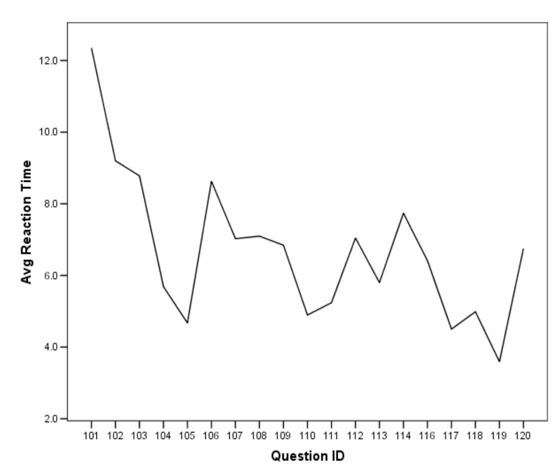

Figure 7. Simple1 Good Forward

As depicted in Figure 8 the response time for the questions decreased as the questions progressed. In both the *simple1 good forward* diagram and *simple2 bad forward* the same downward linear trend occurred. This trend was seen irrespective of the question order being forward or backward.

The *complex* diagram was also graphed to get a visual representation of the average response time per question. However, the same downtrend in response time was not as evident. The *complex* diagram seemed to have a non-linear shape in both *good* and *bad* diagrams as seen in Figure 9 of the *complex bad* diagram.

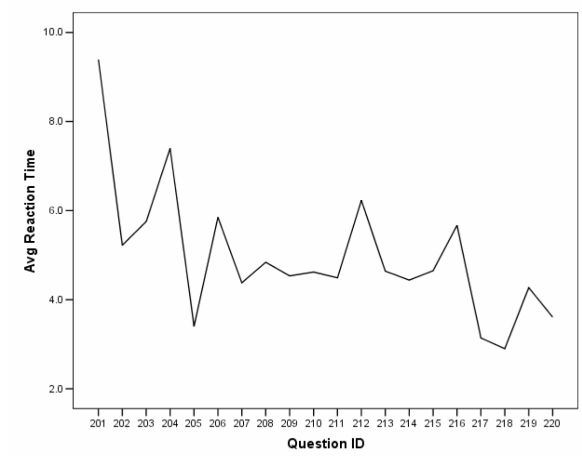

Figure 8. Simple2 Bad Forward

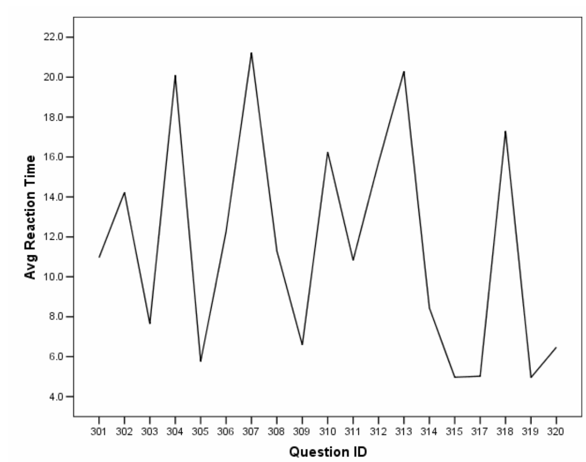

Figure 9. Complex Bad Forward

The questions were then examined on an individual basis to try to explain why this trend occurred in *simple1* and *simple2* diagrams and was not seen in the *complex* diagram. Careful inspection of the questions and diagrams led to a hypothesis that because *simple1* and *simple2* diagrams had fewer connections and boxes, the same boxes had to be used multiple times in questions. Subjects were then able to locate the boxes much faster than if they had to search an entire diagram for different boxes. Also because of the smaller numbers of boxes and connections some boxes were mentioned in adjacent questions which created a chain (i.e. question 3 “Does Response Information send messages to Data Update” question 4 “Does Incident Report send messages to Response Information”) therefore making it less taxing and laborious to find the box and connection.

In the *complex* diagram there were three times as many boxes and connections which meant that the same boxes were not used repeatedly in questions. Also some boxes that were connected were not as close together as they were in *simple1* and *simple2*. In the *complex* diagram it is apparent that sharp peaks occurred for certain questions which sparked more investigation into what other variables besides *good/bad*, *diagram type*, and *question order* that might have affected the comprehension of the subjects.

To examine the data more thoroughly, additional variables were analyzed in order to determine what affected the subjects’ comprehension. The variables were based on assumptions of how subjects examined the diagram and mentally moved from box to box. For example, a student could have traced a line from one box to another. Therefore, the *linelength* variable was examined. Also some of the same variables Purchase studied were analyzed. Eight new variables were studied for these purposes which are listed in Table 2. The variables were used to analyze the diagrams on a per question basis. For example, if a question did not have a connection then the variables *crowflies*, *linelength*, *numbends*, and *crosses* would be zero. If a question did contain a connection between boxes then the appropriate measures were taken. The *boxespd* and *numlinespd* metrics were static numbers for all questions on the same diagram. Before the analysis was done, a new hypothesis was devised.

Hypothesis 3: Higher values of numbends, crosses, and crowflies would contribute to slower response times by subjects.

The analysis consisted of linear regressions using a stepwise method that could determine a predictive model for response time based on the new variables. The threshold was 0.15. The first linear regression model included all questions from each diagram, namely both good and bad versions of *simple1*, *simple2*, and *complex*. In this first model the independent variables that were significant included those in Table 3. The linear regression model Equation (1) relates the post hoc significant variables to the dependent variable response time. The equation contains two variables above the 0.1 significance level that was previously chosen.

$$rt = 33.7 + 2.8\,numlinespd - 6.5\,boxespd + 1.1\,numbends - 0.3\,linesin - 0.5\,crossings \quad (1)$$

Other variables were not significant in the prediction of the dependent variable response time. From these initial results more analysis was done on subsets of diagrams to determine if different variables affected comprehension in different diagrams.

The next linear regression included all *good* diagram questions. The significant variables are listed in the Table 4. Note that none of the line attributes were significant. The Equation (2) gives the significant variables that are related to response time for all *good* diagrams.

$$rt = -38.6 + 9.7\,boxespd - 3.9\,numlinespd \quad (2)$$

Table 2: Post-hoc Variables

Variable name	Variable description
crow flies	distance between the center of two connecting boxes
linelength	length of lines in inches
numbends	number of bends a line contained
crosses	number of lines crossing a single line
linesin	number of lines going into a box
linesout	number of lines leaving a box
numlinespd	number of lines per diagram
boxespd	number of boxes per diagram

The other linear regression was for all *bad* diagram questions. The significant variables are listed in Table 5. Equation (3) gives the significant variables that relate to the response time for all *bad* diagrams.

Table 3: All Diagram Linear Regression Values

Significant variables	*p*-value
numlinespd	0.001
boxespd	0.001
numbends	0.042
linesin	0.106
crossings	0.137

Table 4: Good Diagrams Linear Regression Values

Significant variables	*p*-value
boxespd	0.009
numlinespd	0.011

Table 5: Bad Diagram Linear Regression Values

Significant variables	*p*-value
numlinespd	0.000
boxespd	0.000
linesin	0.079

$$rt = 67.4 + 6.4numlinespd - 14.7boxespd + 0.03linesin \quad (3)$$

Because the same variables were reoccurring in the linear regression models, a principal components analysis was conducted to explain which underlying factors were causing the same variables to always be dominant and determine what or if any of the variables were highly correlated with one another. The stopping rule for the factor analysis was where the eigenvalues were greater than one and the Varimax rotation was also used. Two components were extracted. The values listed in Table 6 give the correlation coefficients between the variables and components.

Component 1 was highly correlated with:

- *linelength*
- *numbends*
- *crossings*

Component 2 was highly correlated with:

- *boxespd*
- *numlinespd*

Component 1 describes the internal dynamics of the graph. It gives an account of what the connections and boxes are in the diagram. Component 2 describes the totality of the diagram. It represents the static values that do not change regardless of the connections present in the diagram. The two components that were extracted from the principal components analysis were then used in another linear regression model to reveal more underlying factors that attributed to the response time. This regression analysis revealed the same results as the principal component analysis.

A correlation model was done to investigate how the variables that were not significant in the linear models were related to the significant variables. This post-hoc analysis result helped to strengthen the conclusions drawn from the earlier preliminary study results. Table 7 shows which variables are highly correlated with one another. The hypothesis was that *crowflies* distance would be one of the variables that would help to predict the subjects' response however it failed to show up in any of the linear regression models. Table 7 shows that the *crowflies* variable is weakly correlated with *boxespd*, *linelength*, and *crossings.*

Table 6: Factor Pattern

	Component	
	1	2
linelength	0.834	-0.295
numbends	0.790	0.124
crossings	0.743	0.047
boxespd	0.182	0.960
numlinespd	0.182	0.960
linesin	0.499	0.102
linesout	0.434	0.067
crowflies	0.339	-0.409

4.4 Discussion of Post-Hoc Results

The post-hoc analysis was used to probe into which attributes contributed to the diagrammatic comprehension of the diagrams presented in this study. Therefore a third hypothesis was proposed to investigate this question (Refer to Hypothesis 3).

In each of the linear regression models for the diagrams, the *crowflies* metric was never significant or a contributing factor to the response time variable and did not show up in any of the linear regression models. The *numbends* variable showed up in the linear regression model that included all diagrams. Other variables such as the static attributes, *boxespd* and *numlinespd*, showed up in all of the linear regression models for all diagrams and in separate models for the good and bad diagrams.

The principal component analysis gave insight and support for previous findings. Two components were extracted using the factor analysis and component 1 included the *linelength*, *numbends*, and *crossing* attributes. Both the *numbends* and *crosses* variables contributed to slower response times. All three of these variables describe the diagram and its internal workings. Component 2 included *boxespd* and *numlinespd* which describe the external portion of the diagram. This supports the fact that the complexity of

Table 7: Correlation of post-hoc variables

	crow-flies	*line-length*	*numbends*	*crossings*	*boxespd*	*numlinespd*	*linesout*	*linesin*
crowflies	-	0.333	0.047	0.144	-0.185	-0.185	-0.023	0.090
linelength		-	0.560	0.442	-0.096	-0.131	0.303	0.360
numbends			-	0.583	0.247	0.221	0.186	0.226
crossings				-	0.173	0.148	0.231	0.191
boxespd					-	0.937	0.098	0.174
numlinepd						-	0.059	0.163
linesout							-	0.107
linesin								-

the diagram does contribute to diagrammatic comprehension.

The principal component provided more insight about which variables were highly correlated with one another. By using the principal component analysis and grouping the variables together, the highly correlated variables such as *boxespd* and *numlinespd* were connected to Component 2. Therefore the variable *crossings* showed up as significantly related variable to component 1.

4.5 Threats to Validity

By using two different question orderings a potential threat caused by the ordering of the questions was assessed. The analysis showed that the direction of question order had no discernible effect on either accuracy or time. By balancing the presentation of good and bad diagrams, the influence of individual abilities was balanced.

Although we had only three pairs of diagrams, decreased response times from *simple1* to *simple2* diagrams suggest that a learning effect occurred between the subjects viewing the first and second diagrams. This threat to validity was not studied further. The increase of time from the *simple2* to the *complex* diagram could correspond to subjects having a higher mental workload because the complexity increases (3 times more boxes and edges than *simple1* and *simple2*).

Due to the fact that the subjects were students, a threat to external validity is present. The diagrams were simpler than software engineers would encounter on a project. The complex diagram tried to mimic real-world diagram complexity, but still lacked certain features present in software-engineering diagrams such as labeling on the edges and the presence of varying arrowheads on the edges, and was relatively simple compared to real project diagrams. The task the subjects performed was similar to a task of software engineering professionals but the working pressure and environment was different than what software engineers would encounter.

5. CONCLUSIONS

This initial experiment investigated the basic question of whether Gestalt principles of proximity, similarity, and continuity affect diagram comprehension. The results of this study showed that the Gestalt principles did affect the comprehension in the complex diagram. Our results from the post-hoc analysis showed that the size of the diagram was a factor. The number of boxes and lines affect the response time of the subjects. Results from the post-hoc analysis gave insight that crossing did have an effect on comprehension. Also, line length and number of bends per line produced slower response times from subjects.

From these results we are able to make assumptions about the proximity of the boxes which is a Gestalt principle that was applied in the good version of the diagrams. By controlling the proximity and placement of boxes on the diagram the line length and number of bends attributes would also have been controlled or reduced. This hypothesis is deduced from results because the majority of the boxes with incoming bended lines were those that had a long distance between them as compared to other boxes that were connected that were near each other. The Gestalt principles of perceptual organization show promise in easing the task of comprehending software-engineering diagrams. This research will help software engineers to determine what type of artistic approaches to consider when designing the software architecture diagrams to help avoid errors in the later stages of the software-engineering lifecycle.

Future work will include investigation of more sophisticated software engineering diagrams that include more boxes and lines and possibly syntax and semantic issues. Work will be done that will investigate other aesthetic criteria in addition to the ones already investigated where each attribute is more strictly controlled and compared to other diagrams where those aspects are not controlled.

6. Acknowledgments

Our thanks go to Ginger Cross, student subjects, MSU ESE research group, Byron Williams, and Chevonne Thomas. This work was supported in part by NSF grant CCR-0132673.

7. References

[1] J. A. Anderson, *Cognitive Psychology and Its Implications*, W. H. Freeman and Company, New York, New York, 1990, 66-68.

[2] M. D. Dodd and J. Pratt, "Allocating Visual Attention to Grouped Objects," *European Journal of Cognitive Psychology*, vol. 17, no. 4, 2005, 481-497.

[3] B. C. Hungerford, A. R. Hevner, and R. W. Collins, "Reviewing Software Diagrams: A Cognitive Study," *IEEE Transactions on Software Engineering*, vol. 30, no. 2, Feb 2004, 84-95.

[4] T. Klemola and J. Rilling, "Modeling Comprehension Processes in Software Development," *Proceedings: First IEEE International Conference on Cognitive Informatics*, Aug 2002, 329-336.

[5] K. Lemon, E. Allen, J. Carver, and G. Bradshaw, "Gestalt Principles Applied to Software Engineering Diagrams: A Preliminary Study." *Proceedings of the 2006 International Symposium on Empirical Software Engineering (Short Papers Track).* Sept. 21-22, 2006, Rio de Janeiro, Brazil, 48-50

[6] J. Reason, *Human Error*, Cambridge University Press, Cambridge, United Kingdom, 1990.

[7] H. C. Purchase, R. F. Cohen, and M. I. James, "An Experimental Study of the Basis for Graph Drawing Algorithms," MIT Press, Cambridge, Massachusetts, 1997.

[8] H. C. Purchase, L. Colypoys, M. McGill, and D. Carrington, "UML Collaboration Diagram Syntax: An Empirical Study of Comprehension," *Proceedings of the First International Workshop on Visualizing Software for Understanding Analysis*, 2002

[9] H. C. Purchase, M. McGill, L. Colypoys, and D. Carringtion, "Graph Drawing Aesthetics and the Comprehension of UML Class Diagrams: An Empirical Study," Australian Symposium on Information Visualization, Sydney, December 2001.

[10] P. Thagard, *Mind: Introduction to Cognitive Science*, MIT Press, Cambridge, Massachusetts, 2005.

[11] C. Ware, H. Purchase, L. Colpoys, and M. McGill, "Cognitive Measurements of Graph Aesthetics," *Proceedings of the First International Workshop on Visualizing Software for Understanding and Analysis*, June 2002, 103-110.