Graph drawing aesthetics and the comprehension of UML class diagrams: an empirical study

Helen C. Purchase, Matthew McGill, Linda Colpoys and David Carrington

School of Information Technology and Electrical Engineering
University of Queensland
St Lucia, Brisbane 4072, Queensland

{hcp, davec}@itee.uq.edu.au

Abstract

Many existing automatic graph layout algorithms are unrelated to any particular semantic domain. Designers of such algorithms tend to conform to layout aesthetics, and claim that by doing so, the resultant diagram is easy to understand. Few algorithms are designed for a specific domain, and there is no guarantee that the aesthetics used for generic layout algorithms will be useful for the visualisation of domain-specific diagrams (for example, visual programs, or entity-relationship diagrams). This paper describes a study which aimed to identify the most important aesthetics for the automatic layout of UML class diagrams from a human comprehension point of view. The results suggest that for specific domains, the actual semantics of the given graph may need to be considered before an appropriate graph drawing can be produced.

Keywords: UML class diagrams, graph layout aesthetics, human performance.

Introduction

CASE tools which provide support for UML diagramming (eg Rational Rose (Rational Rose 2001), Microsoft Visio (Microsoft Visio 2001), Enterprise Architect (Enterprise Architect 2001)) can benefit from the use of an automatic layout tool. Thus, once the user has created a UML diagram, or added new objects and relationships to an existing diagram, a graph layout tool could automatically re-position the objects and lines so that the diagram is more comprehensible.

Many automatic layout algorithms already exist (Battista et al. 1994): they take as input a relational graph structure of objects and the relationships between them, and produce a visual representation of the information in diagrammatic form. The designers of these algorithms tend to optimise certain aesthetic criteria (Coleman and Stott Parker 1996), and claim that by doing so, the resultant graph drawing helps the reader to understand the information embodied in the graph. (These "aesthetic criteria" have been defined and subsequently used in graph layout algorithms by researchers of automatic graph layout algorithms: they do not necessarily relate to the notion of "aesthetically pleasing" with respect to pre-attentive visual perception.) However, these algorithms have typically been defined with respect to abstract graph structure (i.e. nodes and relationships that have no relationship to objects in the real world), and also have not taken any account of human computer interaction issues relating to diagram comprehension.

If CASE tools are to benefit from the use of these automatic layout algorithms, it is important that the most appropriate algorithm, embodying the most appropriate graph layout aesthetic criteria, be chosen to ensure that the diagrams produced are suitable for human comprehension in the intended CASE domain.

Recently, some human experimental work has been performed on the aesthetics underlying common graph drawing algorithms (Purchase 1997): these have shown that the aesthetics of minimising crosses and bends, and maximising symmetry may assist with human performance in graph theoretic tasks on abstract graph drawings. These initial experiments were domain-independent: the graphs used embodied meaningless objects and relationships. There is no guarantee that the results of these domain-independent experiments would necessarily transfer across to the domain of UML diagrams.

Some preliminary work has been done on subjects' preference for different aesthetics in UML class and collaboration diagrams (Purchase et al. 2000), revealing that users preferred diagrams with fewer bends and crosses, shorter edge lengths and an orthogonal structure. However, that experiment only looked at subjects' personal preference for the aesthetics, rather than their performance on UML related tasks.

This paper describes two experiments that aimed to determine which graph drawing aesthetics are most important for the display of UML class diagrams, not with respect to computational efficiency, designers' preference, or even subjects' preference, but with respect to the extent to which the aesthetics produce diagrams that are easy for subjects to understand. The two experiments had identical methodology: the difference between them was in the manner in which the experimental diagrams were produced. In experiment A, aesthetics were measured computationally; in experiment B, they were measured perceptually.

1.1 Experimental aim

The aim of this study was to determine which of the aesthetics underlying common graph drawing algorithms are most suited to human comprehension of UML

diagrams. By asking subjects to perform comprehension tasks on the same UML diagram portrayed with different aesthetic emphases, we aimed to identify the aesthetic criteria that resulted in the best performance. Two experiments were conducted: the first (Experiment A) used computational metrics to determine the presence of different aesthetics in the UML diagrams used; the second (Experiment B) included a preliminary perception experiment which asked for subjects' opinions on the extent of the aesthetics in the diagrams.

1.2 UML class diagrams

UML class diagrams are used to describe the static view of an application (Rumbaugh et al. 1999): the main constituents are classes and their relationships. A class is a description of a concept, and may have attributes and operations associated with it. Classes are represented as rectangles. A relationship between two classes is drawn as a line. Inheritance relationships indicate that attributes and operations of one class (the "superclass") are inherited by other classes (the "subclasses"), without needing to be explicitly represented in the subclasses themselves.

Figure 1 is an example of a small UML class diagram, showing the relationships between the classes in a vehicle hire organisation, including inheritance relationships between the vehicle, car and truck classes.
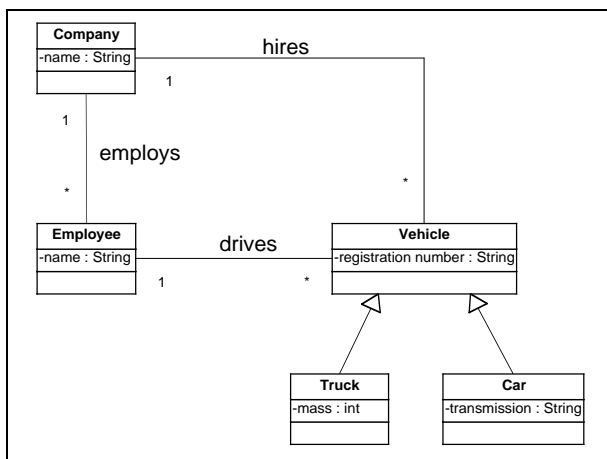

Figure 1: Example UML class diagram.

1.3 Aesthetic criteria

Five graph drawing aesthetics were used in experiment A:

- (b) Minimise bends (the total number of bends in polyline edges should be minimised (Tamassia 1987))

- (n) Node distribution (nodes should be distributed evenly within a bounding box (Coleman and Stott Parker 1996))

- (ev) Edge variation (edge lengths should be uniform (Coleman and Stott Parker 1996))

- (f) Direction of flow (a consistent direction of edge flow (Waddle 2000))

- (o) Orthogonality (fix nodes and edges to an orthogonal grid (Tamassia 1987, Papakostas and Tollis 2000))

A further two aesthetics were included in experiment B:

- (el) Edge lengths (edge lengths should be short; edge lengths should not be too short (Coleman and Stott Parker 1996))

- (s) Symmetry (where possible, a symmetrical view of the graph should be displayed (Eades 1984, Gansner and North 1998))

Experiment A

1.4 Experimental materials:

1.4.1 The application domain

The class diagram used was based on a simple domain, which models a small Information Technology company that employs consultants, programmers and administrative staff to undertake projects for clients. The example includes 13 objects, 12 associations and 5 inheritance relations (see Figure 2).

A textual specification of this domain was produced in simple English. The subjects were asked to match the experimental diagrams against this specification.

1.4.2 UML tutorial and worked example

A tutorial sheet explained the meaning of UML class diagrams, and, using a simple example, described its semantics. Subjects were not expected to have any prior knowledge of UML, and this tutorial provided all the UML background information they required for the experimental task. A worked example demonstrated the task that the subjects were to perform, by presenting a small specification with four different diagrams, and for each diagram indicating whether it matched the given specification or not. Care was taken to ensure that neither the tutorial nor the worked example would bias the subjects towards one layout over another.

1.5 The experimental diagrams

The experimental diagrams were produced according to computational metrics that measured the presence of each aesthetic in a diagram (Purchase 2001). These metrics were scaled to lie between 0 and 1, where 1 means a positive amount (i.e. an amount of the aesthetic for which it is assumed the drawing is easier to read: few bends, high degree of orthogonality, low edge variation, even node distribution, upward flow).
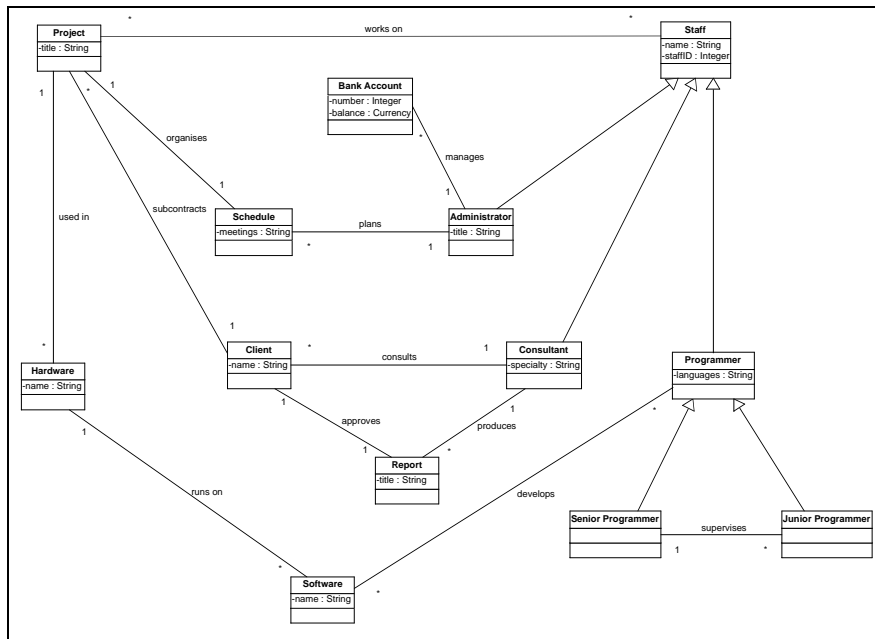
Figure 2: The UML class diagram used for both experiment A and experiment B.

For each aesthetic, a "low-effect" (-) and a "high-effect" (+) version of the diagram was produced.[1] To ensure that there were no confounding factors between the aesthetics, the ranges were controlled as much as possible. For example, to remove any confounding factors in a diagram pair for a particular aesthetic, the measurement of all other aesthetics were kept within a "middle-effect" range. This ensured that any significant difference in the performance of a low-effect diagram with respect to its high-effect counterpart could be attributed to the relevant aesthetic, rather than to any other aesthetic variation within the diagram pair.

Prior work has shown conclusively that edge crossings are an impediment to human comprehension of graph drawings (Purchase et al. 1995, Purchase 1997), so all diagrams had no edge crossings.

A control diagram that conformed to a "middle-effect" range for all the aesthetics as much as possible was also created. There were therefore a total of 11 experimental diagrams.

In addition, a second middle-effect diagram was produced: this was the example diagram that was given to the subjects during the preparation period.

Table 1 shows the aesthetic values for all the diagrams.

Ten incorrect diagrams were created by randomly changing the origin or destination of one relationship per diagram. The layouts of the incorrect diagrams were visually comparable to those of the correct diagrams: as we did not intend to analyse the responses to the incorrect diagrams, their layout was not important. However, it was, of course, important to include incorrect diagrams in the experimental set (so that the correct answer to each diagram presented was not the same), and for these incorrect diagrams to be visually comparable to the correct diagrams (so they could not be identified by mere visual pattern matching).

1.6 Experimental procedure

1.6.1 Preparation

The students were given preparatory materials to read as an introduction to the experiment. These documents consisted of a consent form, an instruction sheet, a tutorial on UML class diagrams and notation, and a worked example of the experimental task. The worked example demonstrated the type of error that had been included in the incorrect diagrams.

As part of this document set, the subjects were also given the textual specification of the UML case study to be used in the experiment: this was the specification against which they would need to match the experimental

---

[1] Note that, because of the way the metrics have been defined, "low-effect" diagrams embody an amount of the aesthetic for which it is assumed the diagram would be difficult to read (for example, many bends, a wide variation in edge lengths), while "high-effect" diagrams embody an amount of the aesthetic for which it is assumed the diagram would be easy to read (for example, a majority of the directed edges pointing upwards, an even distribution of node positioning)

Diagram	Aesthetic				
	bends (b)	orthogonality(o)	edge variation (ev)	node distribution(n)	direction of flow (f)
b+	**1**	0.43	0.66	0.59	0.6
b-	**0.71**	0.46	0.64	0.56	0.6
o+	0.85	**0.70**	0.66	0.56	0.4
o-	0.85	**0.32**	0.64	0.56	0.6
ev+	0.85	0.44	**0.74**	0.59	0.6
ev-	0.85	0.41	**0.55**	0.59	0.6
n+	0.85	0.41	0.66	**0.73**	0.4
n-	0.85	0.48	0.64	**0.45**	0.6
f+	0.85	0.44	0.65	0.59	**1**
f-	0.85	0.46	0.66	0.59	**0**
control	0.85	0.45	0.66	0.57	0.6
example	0.85	0.44	0.66	0.56	0.6

Table 1: The computational aesthetic values for the experiment A diagrams.

diagrams. The subjects were asked to study this specification closely, and memorise it if possible. They were also given an example diagram modelling the specification, with comparable aesthetic metric values to the middle-effect control diagram.

The subjects were given 15 minutes to sign the consent form, read through and understand the materials, ask questions, take notes, or draw diagrams as necessary.

1.6.2 Online task

The subjects then used an online system to perform the experimental task. A copy of the text specification with the example diagram was placed in front of the computer for easy reference, and the UML experimental diagrams were presented in random order for each subject. The subjects gave a yes/no response to each presented diagram, indicating whether they thought the diagram matched the specification or not: two keys on the keyboard were used for this input.

16 practice diagrams (randomly selected from the 21 experimental diagrams) were presented first. The data from these diagrams was not collected, and the subjects were not aware that these diagrams were not part of the experiment. These diagrams gave the subjects an opportunity to practise the task before experimental data was collected.

The 11 correct diagrams were presented twice and the 10 incorrect diagrams once, a total of 32. The diagrams were presented in a different random order for each subject, in blocks of eight, with a rest break between each block (the length of which was controlled by the subject).

Each diagram was displayed until the subject answered Y or N, or 50 seconds had passed. A beep indicated to the subject when the next diagram was displayed after a timeout (which was recorded as an error). The practice diagrams helped the subjects get used to the length of the allocated time period. The timeout period and the time needed for the subjects to prepare for the experiment were determined as appropriate through extensive pilot tests.

A within-subjects analysis was used to reduce any variability that may have been attributed to differences between subjects: thus, each subject's performance on one layout was compared with his or her own performance on an alternative layout. The practice diagrams and the randomisation of the order of presentation of the experimental diagrams for each subject helped counter the learning effect (whereby subjects' performance on the task may improve over time, as they become more competent in the task).

The response time and accuracy of the subjects' responses to the 32 experimental diagrams were recorded by the online system.

1.6.3 Subjects

The 30 subjects were second and third year Computer Science and Information Systems students at the University of Queensland. They were paid $15 for their time, and, as an incentive for them to take the experiment seriously, the best performer was given a CD voucher.

1.7 Results: experiment A

Both the speed and accuracy of each subject's responses were measured, enabling the analysis of two different measures of understanding.
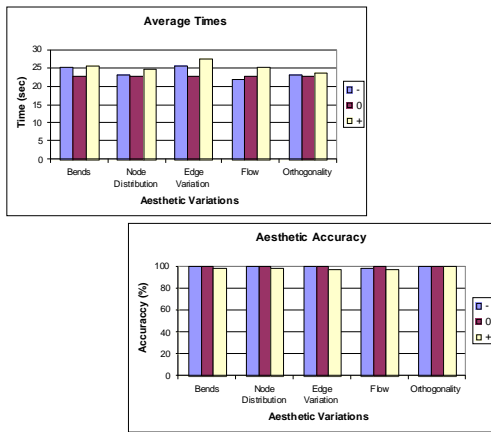
Figure 3: The response time and accuracy results for experiment A.

There were no significant results in the accuracy data: this indicates that the time allocated to the subjects was sufficient for them to correctly classify the diagrams. Thus, only one measurement of understanding was considered - that of the time taken for subjects to respond. Using a two-tailed t-test, the statistically significant response time results are:

Ø Bends

   o control is better than b+ (p<0.05)

Ø Edge variation

   o control is better than ev+ (p<0.05)

   o control is better than ev- (p<0.05)

Ø Flow

   o control is better than f+ (p=0.058, approaches significance)

   o f- is better than f+ (p<0.05)

1.8 Analysis

1.8.1 Bends

The data show that the diagram with a low number of bends (b+) produced worse performance than the control diagram (which had a medium number of bends). This is a surprising result, as a previous study showed that in a domain-independent context, performance is improved with fewer bends (Purchase 1997), and a UML preference experiment showed that subjects did not like bends (Purchase et al. 2000).

A possible explanation for this result may be that increased orthogonality requires an increase in the number of bends, and therefore the diagram with a medium number of bends may have produced a good performance because of an increase in orthogonality. However, the orthogonality values for these two diagrams are not substantially different: 0.43 for b+, and 0.45 for the control. In addition, the lack of any significant results for the orthogonality diagrams o+ and o- implies that increased orthogonality cannot be used as an explanation for this surprising result.

1.8.2 Edge variation

The control diagram (with a medium variation of edge lengths) produced better performance than both ev+ (all edges of similar length) and ev- (some edges very short, some edges very long). This was another surprising result, as we had expected that ev+ would produce better performance than both the control and ev-.

It appears that widely varying edge lengths is less useful than a medium variation of edge lengths: this is as expected. The improved performance of the control over the diagram with edges of similar size is difficult to explain, and led us to believe that perhaps it is the actual length of the edges (rather than their variation) that may be important.

1.8.3 Flow

Both the results for the flow diagrams show that there was decreased performance on the diagram with the majority of the edges directed upwards (f+). Again, this result is contrary to expectations. A study of UML class diagram syntax (Purchase et al. 2001) showed an improved performance, and an increased preference, for upward arrows, as it is more intuitive to have the superclass placed above the subclasses. As the f+ and f- diagrams were almost mirror images of each other (about a horizontal axis), there were no obvious confounding factors that produced this unexpected result.

1.9 Discussion

None of our expectations were satisfied in experiment A: two of the aesthetics (node distribution and orthogonality) produced no significant results at all, and the significant data from the other three aesthetics was difficult to interpret reasonably and consistently.

In reassessing the diagrams that we used for this experiment, we felt that perhaps the problem was in the measurement of the presence of the aesthetics. The metrics, while useful for measuring the aesthetics from a computational point of view, may be less useful for measuring perceptual aesthetic presence from a human point of view. For example, the orthogonality metric measures the extent to which the nodes and edges are placed along an underlying unit grid, but the human perception of orthogonality in a diagram may not match the numerical value produced by the metric. This phenomenon may particularly be the case for aesthetics which are global; that is, require an overall assessment of the entire diagram, for example, orthogonality, symmetry, or node distribution.

We therefore decided to run the experiment again, but this time with a different set of diagrams. The diagrams for experiment B would be created according to humans' perception of the presence of each aesthetic in the diagrams, rather than according to the defined metrics.

Experiment B

1.10 Experimental materials:

The application domain, the UML tutorial and worked example, the preparation period, the online task and the data collection method were all the same as for experiment A. The only change to the experimental procedure was that the timeout was 40 seconds (rather than 50 seconds): this change was due to the fact that as the diagrams for experiment B were produced according to human perception, rather than according to computational metrics, they appeared to the subjects to be easier to read. This timeout period was determined as appropriate through extensive pilot tests. The subject pool for experiment B was the same as experiment A: there were a total of 35 subjects for experiment B.

1.11 The experimental diagrams

The main difference between experiment A and experiment B was the way in which the experimental diagrams were produced. While experiment A used computational metrics to determine the presence of an aesthetic in a diagram, in experiment B, a separate human perception study was used to assess the extent to which aesthetics were perceived in a diagram.

Experiment B differed from Experiment A in two other important aspects: choice of aesthetics and aesthetic variation.

1.12 Choice of aesthetics

Experiment B examined those aesthetics that were tested in experiment A as well as two new aesthetics that it was felt may also have an influence on performance. These two aesthetics were:

Edge lengths. For experiment A, we only considered the variation of the edge lengths. Having got results that seemed to indicate that a medium-effect edge variation (i.e. a variation in the lengths of the edges which is neither small nor large) produces better performance, we decided to include edge lengths in experiment B (Coleman and Stott Parker 1996).

Symmetry. With respect to graph layout, symmetry is best considered perceptually rather than computationally. A computational definition of symmetry which merely considers the geometric correspondence of nodes around vertical and horizontal axes does not take into account local symmetries, and the tolerance that humans have for perceiving symmetry (i.e. the fact that humans will state that a square is symmetric even if the pixel values of the corners are slightly removed from the underlying grid). A computational algorithm that takes all local symmetries and perceptual tolerance into account would be computationally complex, and can only be a very rough model of the human perception of symmetry. It was therefore infeasible to include symmetry in experiment A, when the aesthetics were measured computationally. As the diagrams used in experiment B were created through interviews with humans on their perception of the diagrams, it was more appropriate to include symmetry in this second experiment.

1.13 Aesthetic variation

In experiment A, a single control diagram served as a middle-effect diagram for all the aesthetics. For experiment B, a different middle-effect diagram was produced for each aesthetic. As the analysis was to be done with respect to the variations within the aesthetics, it was not necessary to use the same middle-effect diagram for all the aesthetics. In experiment A, we did so because it was convenient: it was not necessary in experiment B. Thus, for each aesthetic, three diagrams were created by hand: low-effect (-), middle-effect (0) and high-effect (+).

To confirm that these diagrams had an appropriate amount of low-, middle- and high-effect of the aesthetics, and that the aesthetics were appropriately controlled, simple perception experiments were performed with 10 subjects. These subjects who took part in these perception tests were from a comparable subject pool to those who participated in the main experiment.

The subjects were asked to rank sets of three diagrams according to the presence of the aesthetic. For example, a subject was shown the n+, n0 and n- diagrams and asked to rank them according to the extent of even node distribution in the diagrams.

In experiment A, we were able to use the computational metrics to ensure that there were no possible confounds in the diagrams. In experiment B, the possible confounds of symmetry and orthogonality were also addressed in the interviews. For example, the subjects were asked to rank the n+, n0 and n- diagrams according to symmetry, the desired result being that they would find it difficult to do so. We needed to ensure that a difference in performance on the node distribution diagrams could not be attributed to differences in symmetry and othogonality.

The bends and flow aesthetics were not perceptually tested in the production of the diagrams, as their presence is better assessed computationally (for example, by counting the number of bends or counting the number of edges pointing upwards). However, the bends and flow diagrams were tested for the possible symmetry and orthogonality confounds.

A total of 10 incorrect diagrams were created by randomly changing the origin or destination of one relationship per diagram. The layouts of the false diagrams were visually comparable to those of the correct diagrams: as we did not intend to analyse the responses to the incorrect diagrams, their layout was not important. However, it was, of course, important to include incorrect diagrams in the experimental set (so that the correct answer to each diagram presented was not the same), and for these incorrect diagrams to be visually comparable to the correct diagrams (so they could not be identified by mere visual pattern matching).

The 21 correct and 10 incorrect diagrams were each presented once in the online task: a total of 31 experimental diagrams.

1.14 Results: experiment B

Both the speed and accuracy of the subject's response were measured, enabling the analysis of two different measures of understanding.
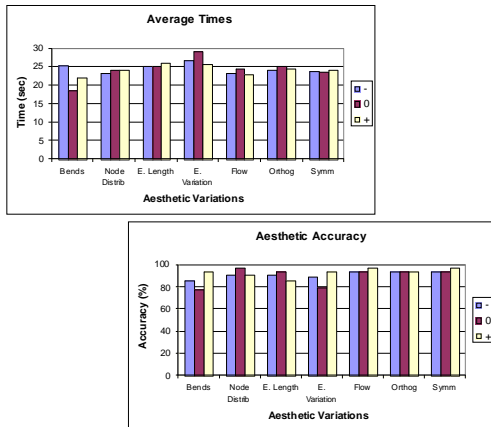

Figure 4: The response time and accuracy results for experiment B.

Unlike experiment A, some significant accuracy data was obtained. This was probably because of the reduced timeout duration (40s rather than 50s), which resulted in more errors.

Using a two-tailed t-test, the statistically significant results are:

Ø Bends

- o b0 is faster than b- ($p < 0.05$)
- o b+ is faster than b- ($p < 0.05$)
- o b+ is more accurate than b0 ($p = 0.057$, approaches significance)

Ø Edge Variation

- o ev+ is faster than ev0 ($p < 0.05$)
- o ev- is faster than ev0 ($p < 0.05$)
- o ev+ is more accurate than ev0 ($p < 0.05$)

1.15 Analysis

1.15.1 Bends

The results for the bends diagram suggest that a reduced number of bends produces the best performance. The accuracy result (that the diagram with least number of bends, b+, is more accurate than the middle-effect diagram, b0), only approaches significance at the 0.05 level. This result conforms to our prediction, and previous studies (Purchase 1997, Purchase et al. 2000).

1.15.2 Edge variation

The data show that the middle-effect edge variation diagram had worse performance than both the diagram with similar length edges (ev+) and the diagram with edges of greatly varying lengths (ev-) - a result contrary to that of experiment A, when the control diagram had the best performance.

These conflicting edge variation results suggest that there are other factors to be considered, including the fact that we obtained no significant results from the diagrams embodying the edge length or node distribution aesthetics.

In the diagrams used in these experiments, no attempt was made to conform to any semantic grouping; thus the nodes were arbitrarily placed in the diagram. It appears that the length of the edges and the spread of the nodes does not matter with such positioning. However, it is possible that performance would be improved if the nodes were not arbitrarily positioned. For example, if the edges and nodes were positioned in a manner that placed semantically related nodes close to each other (even if they are not explicitly joined by an edge), performance could be affected.

1.16 Discussion

Despite our efforts to use diagrams that conformed to the human perception of aesthetics, rather than a computational measure, only one of our expectations (with respect to bends) was satisfied in experiment B: five of the aesthetics (node distribution, edge length, symmetry, flow and orthogonality) produced no significant results at all, and the significant data from the edge variation aesthetic was difficult to interpret without considering the possible effects of the semantics of the diagram layout.

Conclusions

Having attempted two versions of this experiment, and obtained few concrete results, it is tempting to say that none of the aesthetics really matter (apart from bends, which only matters a little), and therefore there would be no human comprehension differences between two UML support tools that use automatic layout algorithms embodying different aesthetics.

We believe that there are additional semantic issues that need to be considered when a layout algorithm is used in a domain-specific tool.

Automatic graph layout algorithms typically do not take the semantics of the diagram into account. As we wished our results to relate to the design of such algorithms, we did not consider the semantics of the diagrams when we created them according to the layout aesthetics.

Our results suggest that improved performance is not merely related to even node distribution, edge lengths or variation of the edge lengths, but requires something else: we suggest that the extra feature that needs to be considered is the semantic grouping of related objects. Even the surprising results for the bends aesthetics could be explained by a break down in semantic grouping that may result from eliminating bends entirely: for example, it may be preferable to add some bends to the diagram if it means that the subclasses in an inheritance hierarchy can be positioned close to each other.

This speculation is based on two sources. First, the Cognitive Dimensions framework proposed by Green and Petre (1996) includes the dimension of "Secondary

Notation." which is defined as "valuable layout cues [that] are typically not formally part of the notation ... but ... can be used to exhibit relationships and structures that might otherwise be less accessible" (Petre 1995). The visual proximity of objects is a secondary notation: Petre (1995) found that placing unrelated objects next to each other gave the misleading impression that they were semantically related. Second, in informal discussions with the subjects, many of them commented that the grouping of semantically related classes was an important layout feature.

Further studies could attempt to validate this idea. We can envisage a similar experiment to the ones described in this paper, but with the diagrams produced according to varying levels of semantic grouping. Such an experiment could help determine the extent to which semantic grouping is necessary for improved human comprehension.

Another interesting informal comment from the subjects was related to the nature of the task and the form of the experimental materials. Students said that they found the diagrams easier to understand if, when reading from top to bottom, the order of the classes matched their order in the given written specification.

This comment demonstrates one of the limitations of this experiment. Any formal empirical study has limitations: in our case, we were using university students as subjects, rather than software engineers, and the comprehension task and application were constrained to a simple domain and matching task. We chose the task of noticing associations for which the source or destination was incorrect as one way of measuring the comprehension of the diagram: there are many other ways in which comprehension may be assessed, especially in relation to a real-world application task. More extensive case studies that follow the use of UML in an industrial application, or that observe the use of UML support tools in practice would give a greater insight into suitability of the different aesthetics and the importance of semantic grouping, from a human comprehension point of view.

In choosing a graph layout algorithm to use in a CASE tool, its suitability for comprehension needs to be considered. While different generic algorithms, embodying a variety of aesthetics, may produce diagrams that look attractive, a "nice" layout is unlikely to be sufficient for intuitive use. Algorithms that have been specifically designed for UML, and which are able to take into account the semantics of the diagram, are more likely to be effective from a human understanding point of view.

Acknowledgements

References

COLEMAN, M. and STOTT PARKER, D. (1996): Aesthetics-based graph layout for human consumption. *Software — Practice and Experience* **26**(12):1415-1438.

DI BATTISTA, G., EADES, P., TAMASSIA, I. and TOLLIS, I. (1994): Algorithms for drawing graphs: An annotated bibliography. *Computational Geometry: Theory and Applications* **4**:235-282.

ENTERPRISE ARCHITECHT (2001) http://www.sparxsystems.com.au/ea.htm, 23 October 2001.

GANSNER, E., and NORTH, D. (1998): Improved force-directed layouts. *Proceedings of the Graph Drawing Symposium 1998.* Montreal, Canada, 364-373, Springer-Verlag.

GREEN, T. and PETRE, M. (1996): Usability analysis of visual programming environments: A ʽcognitive dimensionsʼframework. *Journal of Visual Languages and Computing* **7**:131-174.

MICROSOFT VISIO (2001) http://www.microsoft.com/office/visio/, 23 October 2001.

PAPAKOSTAS, A. and TOLLIS, I. (2000): Efficient orthogonal drawings of high degree graphs. *Algorithmica* **26**(1):100-125.

PETRE, M. (1995): Why looking isnʼt always seeing. Readership skills and graphical programming. *Communications of the ACM* **38**(6):33-44.

PURCHASE, H. (1997): Which aesthetic has the greatest effect on human understanding? *Proceedings of the Graph Drawing Symposium 1997*, Rome, Italy, 248-261, Springer-Verlag.

PURCHASE, H. (2002): Graph drawing aesthetics metrics. *Journal of Visual Languages and Computing* **to appear**.

PURCHASE, H., ALLDER, J. and CARRINGTON, D. (2000): User preference of graph layout aesthetics: A UML study. *Proceedings of the Graph Drawing Symposium 2000*, Colonial Williamsburg, USA, 5-18, Springer-Verlag.

PURCHASE, H., COHEN, R. and JAMES, M. (1995): Validating graph drawing aesthetics. *Proceedings of the Graph Drawing Symposium 1995*, Passau, Germany, 435-446, Springer-Verlag.

PURCHASE, H., COLPOYS, L. and MCGILL, M. (2001): UML class diagram syntax: An empirical study of comprehension. *Proceedings of the Australian Symposium on Information Visualisation*, Sydney, Australia, **this volume**.

RATIONAL ROSE (2001) http://www.rational.com/products/rose/index.jsp, 23 October 2001.

RUMBAUGH, J. JACOBSON, I. and BOOCH, G. (1999): *The Unified Modeling Language Reference Manual*. Reading, Mass, Addison Wesley Longman Inc.

TAMASSIA, A. (1987): On embedding a graph in the grid with the minimum number of bends. *SIAM J. Computing* **16**(3):421-444.

WADDLE, V. (2000): Graph layout for displaying data structures. *Proceedings of the Graph Drawing Symposium 2000*, Colonial Williamsburg, USA, 241-252, Springer-Verlag.