Evaluating the Comprehension of Euler Diagrams

Florence Benoy and Peter Rodgers

University of Kent

Abstract

We describe an empirical investigation into layout criteria that can help with the comprehension of Euler diagrams. Euler diagrams are used to represent set inclusion in applications such as teaching set theory, database querying, software engineering, filing system organisation and bio-informatics. Research in automatically laying out Euler diagrams for use with these applications is at an early stage, and our work attempts to aid this research by informing layout designers about the importance of various Euler diagram aesthetic criteria. The three criteria under investigation were: contour jaggedness, zone area inequality and edge closeness. Subjects were asked to interpret diagrams with different combinations of levels for each of the criteria. Results for this investigation indicate that, within the parameters of the study, all three criteria are important for understanding Euler diagrams and we have a preliminary indication of the ordering of their importance.

Keywords

Euler diagrams, graph drawing.

1: INTRODUCTION

Euler diagrams, represented by interlinking sets, are becoming a widely used technique in information visualization. Often, items shown as dots or graphs are present in the diagram to indicate which sets particular items belong to. A simple Euler diagram is shown in Figure 1. The circular lines, called contours, intersect. The separate areas produced by the intersections are called zones. In this paper we label the contours. The zones can be described by the contours in which they are contained. In this paper we do not consider the outside zone not contained by any contours. One of the main advantages of Euler diagrams is that they allow the visualization of n-ary relationships, using containment and intersection in a mathematically rigorous framework. As a result they represent richer concepts than alternative visualization systems, such as graphs, which are restricted to binary relationships.

It should be noted that these diagrams are often inaccurately called Venn diagrams. In fact, Venn diagrams are a special case of Euler diagrams, where every possible zone is present. The diagram in Figure 1 is not a Venn diagram because the zones **ac** and **abc** are not present, whereas the diagram in Figure 2 is a Venn diagram.

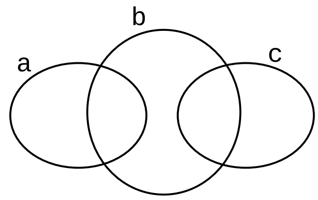

Figure 1. An Euler Diagram with contours: a, b **and** c**; and zones:** a b c ab bc**.**

The automated layout of any kind of diagram carries with it the problem of discerning the criteria for the layout that will most effectively allow the user to interpret the diagram in the intended way. For example, with graphs, certain features such as line crossings are known to have an inhibiting effect on comprehension and consequently most algorithms have metrics that allow them to reduce the number of crossings as far as possible. However, the effect of other criteria and the possible interactions between criteria in particular contexts is less well understood and in practice may be based on cognitive theory and intuition on the part of the researchers and their colleagues. Studies that seek to validate (or otherwise) commonly used criteria by empirical investigation have been pioneered by Purchase [11,13].

The automatic layout of Euler diagrams has only recently been investigated. A multicriteria optimizing system was developed [6], which attempts to improve several metrics, each of which represents an aesthetic feature of the Euler diagram. However, the initial choice of both metrics and the notion of optimal in connection with each of the metrics, was ad hoc and the method employed defining the relative weights assigned to them was not rigorous. This paper starts the process of putting the use of such criteria on a more scientific footing by describing an empirical investigation that compares the effectiveness of metrics for laying out Euler diagrams.

A challenging aspect of Euler diagram layout is embedding the diagram, that is, going from an abstract representation, where just the set intersections are known to a diagram with a layout in 2 dimensional space, so visualizing the set intersections. Generating such an embedding is not a fully solved problem. Flower and Howse [5] implement a mechanism for embedding Euler diagrams under strong wellformedness conditions, where the contours must only meet transversely without sharing line segments and only two contours may meet at any point. As the wellformedness conditions are relaxed, more diagrams can be drawn [4,12]. These embeddings are not typically aesthetically pleasing, and hence work has been performed in improving the layout of embedded diagrams [6]. This work was extended to Euler diagrams enhanced with graphs [10].

The research question we address is the confirmation or otherwise that three particular criteria really do facilitate the comprehension of Euler diagrams and, if possible, to infer an ordering of importance on these criteria. The investigation described here relates specifically to Euler diagrams but the associations are abstract. It could be argued that with only abstract associations the possibility of implicit associations from life experience coming into play is reduced. The choice of criteria under investigation was directed by the metrics used by the multicriteria optimizer of [6]. Several of these metrics were amalgamated into three criteria: Contour Jaggedness, Zone Area Inequality and Edge Closeness. The outcome of the investigation suggests that all three of these criteria, particularly Contour Jaggedness and Edge Closeness have an important impact on comprehension. However, it is apparent that as the diagrams become more complex, interactions between these criteria come into play and further investigation is warranted.

Euler diagrams are used in a wide variety of applications such as visualizing biological data (see Figure 2), visual file system organisation (see Figure 3) and database queries (see Figure 4). Further applications include the familiar use of these diagrams to teach set theory in school, as well as software engineering diagrams [7] and visualizing the overlap of several diagnostic tests.

Euler diagrams are closely related to extended graph notations such as hypergraphs and higraphs [8,9]. These are graphs with some notion of containment and are widely used. For example, several UML diagram types use variants of graph notation extended with containment.

There are widely available software tools for visualizing Euler diagrams. Microsoft PowerPoint 2003 includes a 'Venn diagram' option as one of its six built in diagram types; in fact, this option produces Euler diagrams when the number of circles increases beyond three. The site venndiagrams.com has a database of over 10,000 diagrams created by users of its online application, which visualizes 3-set Venn diagrams.

The remainder of the paper is organized as follows: Section 2 gives the design of the investigation; Section 3 details the results, and our interpretation of the data. Section 4 summarizes the paper and gives some further research directions.

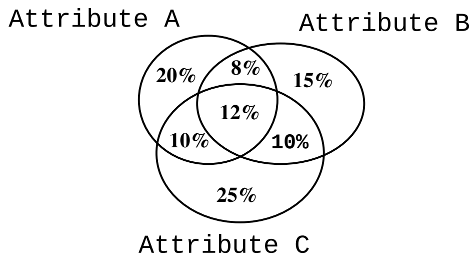

Figure 2. A visualization of biological data. The areas of the zones relates to the percentages of proteins having particular attributes.

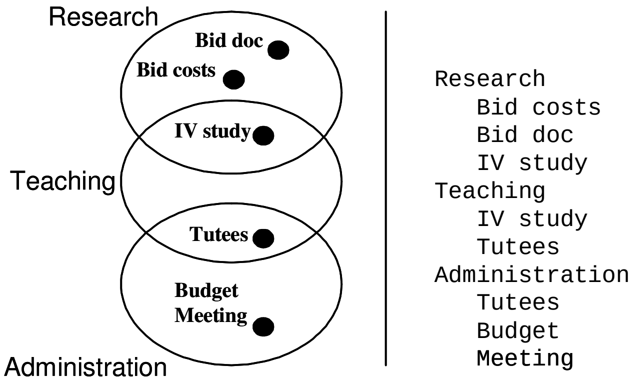

Figure 3. On the left is a Euler diagram based on VennFS [3], showing an Eueler based visualization of a file system. Here, unlike traditional hierarchical file views, shown on the right, documents can be in more than one folder. The 'IV study' file relates to both research and teaching, and can be placed in both categories. Similarly, the 'tutees' file is in two categories. In a hierarchical view either one of the categories must be chosen, or the files must duplicated.

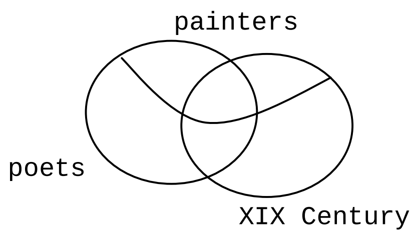

Figure 4. Based on [12] An Euler diagram representing part of a multimedia database. Here, the diagram indicates that there is only data on painters that are poets or alive in the XIX century, and there is no data for painters that are not also in either category.

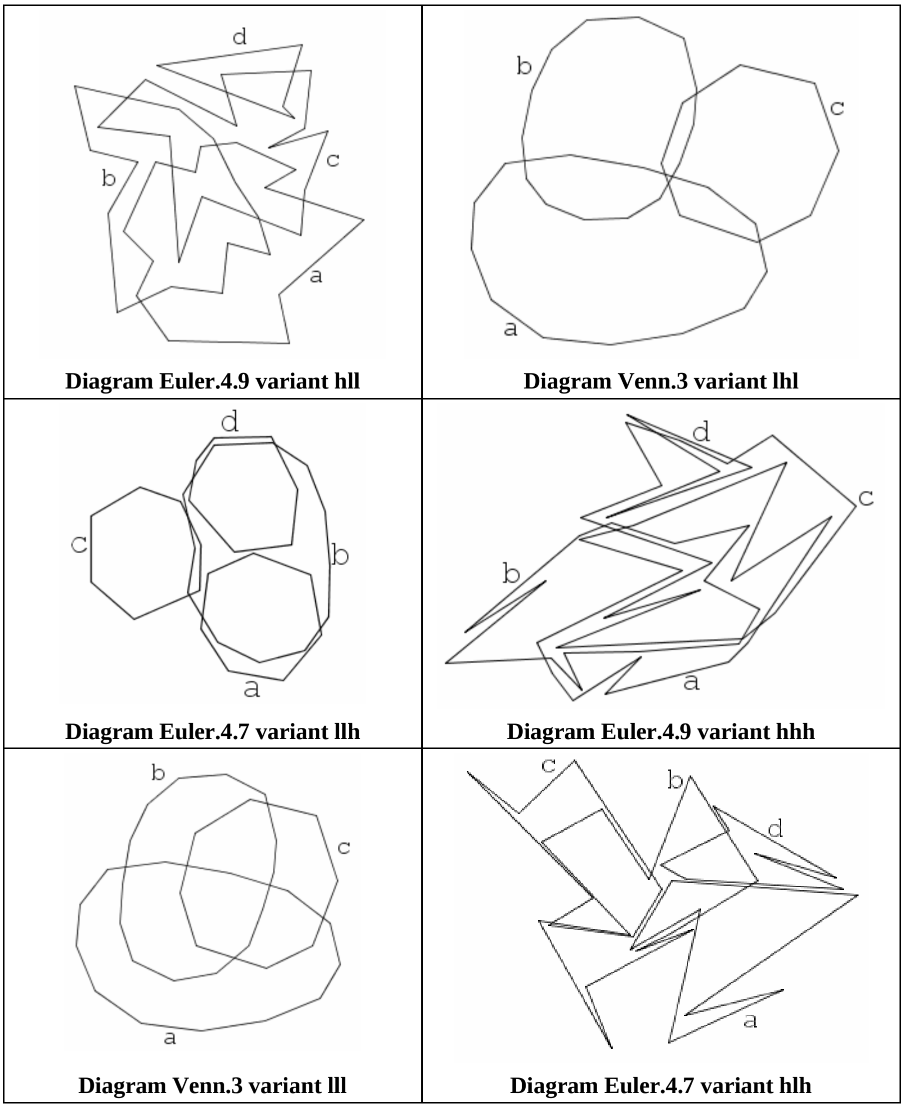

Diagram Euler.4.9 variant hll | **Diagram Venn.3 variant lhl**

Diagram Euler.4.7 variant llh | **Diagram Euler.4.9 variant hhh**

Diagram Venn.3 variant lll | **Diagram Euler.4.7 variant hlh**

Figure 5. Some example diagrams.

2: EXPERIMENTAL DESIGN

The study was designed to gauge the effect of certain layout criteria on the capacity of users to interpret Euler diagrams correctly. To do this we tested the ability of subjects to find the zones in Euler diagrams. The measurements for the metrics were complex and in some cases a wide range of measurements qualified for what was deemed to be a good layout. We labelled each diagram with a level of *low* or *high* according to each of the 3 criteria. The actual numeric values could not be compared across criteria as the ranges differed substantially. The subjective viewpoint held by the implementers of the metrics was that despite these different ranges over the three metrics, overall, the lower the metric, the more likely the outcome was to produce a diagram layout that was good. Each diagram had *low* or *high* levels for each of the three criteria and each different combination of levels is known as a ***variant*** of a particular diagram. Hence, each diagram had 8 variants in all.

2.1: The Criteria

The criteria we chose to investigate were based on our experience with automatically laying out Euler diagrams. For a complete block design, the number of diagrams that need to be shown to subjects increases exponentially with the number of criteria. For this reason and because this was a preliminary investigation of this type, we chose to keep the design as simple as possible whilst still being useful and we limited the study to just three criteria. We chose the three that seemed most likely to facilitate comprehension. This choice was based on the findings of research into graph drawing aesthetics and the experiences of the hill climbing optimizer team.

The number of contours and zones was limited by the actual size of the screen display and by our estimation of what we might reasonably expect our subjects to cope with, given the task. The complexity range of the diagrams shown in the main study was informed by what we learned in the

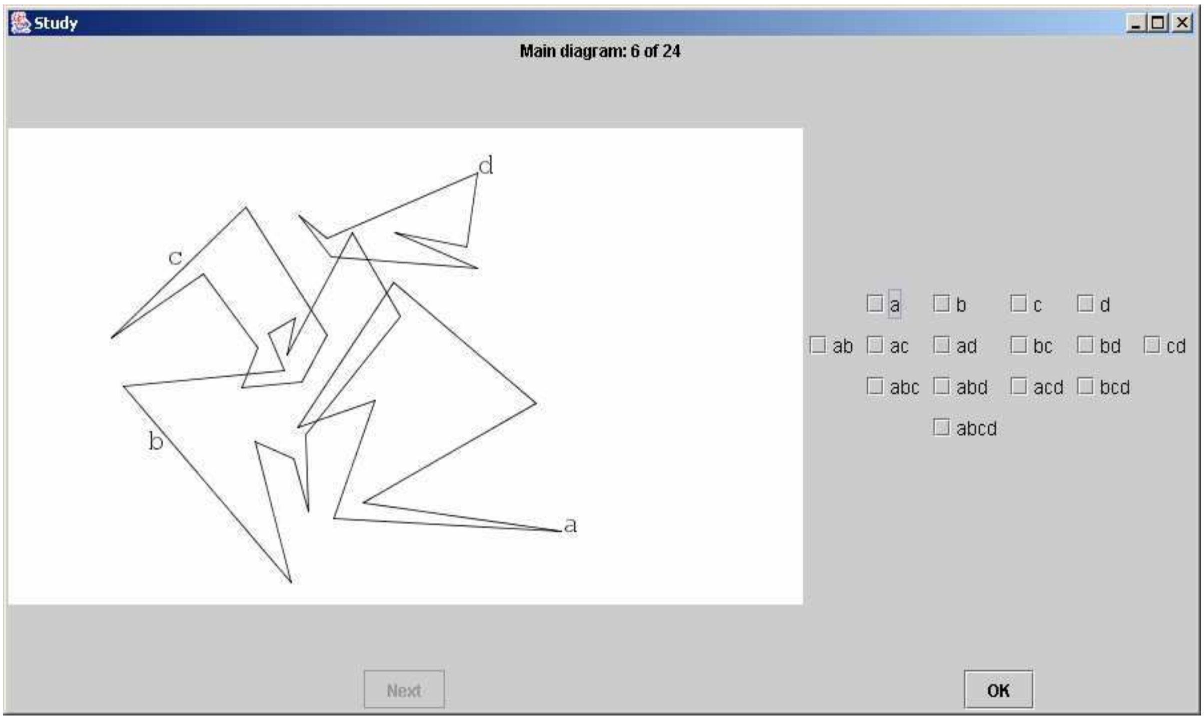

Figure 6. This is a screenshot of the software with diagram Euler.4.7, variant hhl. The zones in this diagram are: a b c d ab bc bd**.**

pilot studies, for example, we withdrew one of our initial set of diagrams altogether, as too complex, see The Experimental Methodology on page 5. Figure 5 gives some examples of the diagrams presented to the subjects. There were three different logical diagrams: Euler.4.9, a four contour Euler diagram with 9 zones; Euler.4.7, a four contour Euler diagram with 7 zones; and Venn.3, the Venn diagram with 3 contours.

The chosen criteria were: *Contour Jaggedness*, *Zone Area Inequality* and *Edge Closeness*. When describing the diagram variants, the convention we use is to take the criteria in the order: Contour Jaggedness, Zone Area Inequality and Edge Closeness, so that **hhh** is *high* for all the criteria, whereas, **hlh** is *high* for Contour Jaggedness and Edge Closeness, but *low* for the middle criterion, Zone Area Inequality.

The criteria in detail are:

- Contour Jaggedness relates to the continuousness of the contour lines. This means that smooth lines would have a relatively *low* measurement and jagged lines a *high* one. The diagram on the top left of Figure 5 has a *high* Contour Jaggedness measurement, but is rated *low* for the other two criteria.
- Zone Area Inequality relates to the relative sizes of the zone areas. An uneven distribution, with some zones very large and some very small will have a *high* measurement, whereas an even distribution with all zones closer in size will have a *low* one. The diagram on the top right of Figure 5 has a *high* Zone Area Inequality level, but is *low* in the other two criteria.
- Edge Closeness relates to the closeness of lines from different contours. Diagrams with lines close together for large sections have *high* measurements, diagrams with lines always diverging will have *low* ones. The diagram on the middle left of Figure 5 has a *high* Edge Closeness level, but is *low* in the other two criteria.

2.2: Generation of the Test Diagrams

The starting point for all of the diagrams was generated using the diagram layout method described in [6] with the settings that had been assessed as the most effective. The effectiveness of those settings was based on the visual perception of the researchers. The quality metrics for each diagram were recorded and then the diagram was adjusted by hand in order to toggle one or two of the attributes from the initial *low* measure to a *high* one. Maintaining uniform zone areas whilst toggling the other two attributes of the diagrams was not straightforward and the acceptable range of numeric values for both the *low* and *high* measures of Zone Area Inequality were uniformly wider than for the other two criteria.

2.3: Software

The study required software that could display an Euler diagram, take as input the zones the subject thinks are present in the diagram, and output the results for all of the diagrams at the end of the session. Figure 6 shows a screen shot of the system in operation. The check boxes on the right correspond to all possible zones for the given contours. The subject then checks the boxes corresponding to the zones that he thinks are present in the diagram. After clicking "OK" for a diagram, the diagram was removed and the timing was paused, allowing subjects to take a rest, if they wished. The subject clicks the "Next" button to move on to the next test. After all the diagrams were presented to the subject, the results were displayed in a scrolling window containing all the diagrams, the subject's answer and the correct answer.

It was considered that logging a subject's responses using the check box in the way outlined above was less prone to accidental error than requiring the subjects to type in their solutions. Also, having a list of possible zones would further reduce the possibility of typing errors. It is inevitable that for the more complex diagrams subjects will develop strategies for finding solutions that will vary both between and within the individual subjects. By including the subject in our statistical models we hoped to take account of this effect as far as possible.

2.4: The Experimental Methodology

The study consists of subjects attempting to choose the correct zones for each of a sequence of Euler diagrams. For the main study we had 3 different diagrams and 8 combinations of the three criteria: Contour Jaggedness, Zone Area Inequality and Edge Closeness. This gives a total of 24 main diagrams, some of which are shown in Figure 5. The subjects were given one of 24 randomized sequences of diagrams, this number is coincidentally the same as the number of main diagrams. At the beginning of the session the subjects were asked to read through a handout explaining the requirements of trials. This was accompanied by a verbal introduction to the material in the handout and a by a demonstration of the task and the opportunity to ask questions. Before the main set of diagrams the subjects were given 8 training diagrams, each of which was immediately followed by feedback on their performance and the correct solution. At the end of the session the subjects were given their results, in the form of a screen display that showed them all the diagrams they had been tested on, and an indication of how they performed on each. They were then asked to fill in a questionnaire and given a debriefing document explaining the nature of the study.

We carried out two pilot studies to check our methodology. The first was with six postgraduate students in the Computing Laboratory. As in the main study, we paid the students £5 for attending and a further £5 for a high score, in order to motivate their performance. However, for the main tests an additional prize of £10 was awarded to the subject who performed the best. The subjects were told that this prize would be awarded to the subject with the most accurate result, using time as a tie-break. The first pilot went well with all but one student scoring highly, and all finishing within 45 minutes. We were concerned that subjects may have become too familiar with the study investigations, as the majority of them had taken part in previous pilot studies [1,2], therefore we conducted another pilot study using contacts at the University of Brighton, where we had eight subjects. Again, these subjects were in the main postgraduates, but also included two members of staff. This time two subjects experienced real difficulties, with low numbers of correct solutions and taking well over an hour to finish. Two subjects indicated in the questionnaire that the tests had a high difficulty level. Consequently we reduced the number of main test diagrams to 24 from 32 by removing the 8 variants of one diagram that results from the pilot indicated were the most difficult. Whilst at Brighton, we invited members of staff, experienced in empirical studies to observe the trials, and comment. They indicated general satisfaction with the methodology; however they did suggest that we reduce the high score threshold, so that more students in the main study would reach the threshold. This would support the idea that the additional payment was simply an incentive to take the tasks seriously as the purpose of the study was to evaluate the layout criteria not the subjects.

The subjects for our main study were computing undergraduates. We used computing students both because they were the most accessible, and because they have some knowledge of set theory which is taught using similar diagrams during the first year of study. Hence, the subjects would not require an introduction to Euler diagrams, only to the problem specification and the environment. Also these students could well be representative of the population from which practitioners who might use or be involved in the automated generation of Euler diagrams might be drawn. Each task by each subject was monitored in two ways: i) the time taken to complete the task and ii) whether the task was successfully completed or not. A task was successfully completed when all of the zones present in the diagram and only those zones present in the diagram had been ticked.

3: RESULTS

This section is in three parts, first, the data is summarised in two bar charts, second the statistical results are presented, and finally an interpretation and discussion of these outcomes is given.

For the purpose of this analysis, recall:

There are 3 distinct diagrams Venn.3 - the complete Venn diagram with 3 contours and 7 zones, Euler.4.7 – an Euler diagram with 4 contours and 6 zones and Euler.4.9 – an Euler diagram with 4 contours and 9 zones.

variant – a diagram variant is distinguishable by the ratings of high or low associated with the 3 qualities:

- contour jaggedness
- zone area inequality
- edge closeness

There were 3 distinct diagrams, with 8 variants of each so 24 tasks in all.

correct – a subject either **successfully** identifies all of the zones in the diagram in which case he succeeds or he does not, in which case he fails. We have not considered the results for partially correct failures as the reasons for not identifying the zones correctly could be very diverse, for example a distraction. Our analysis simply focuses on the correct solutions.

There were 49 subjects taking part in the study.

3.1: Summary of the Data

A subject's ***score*** is the number of tasks completed **successfully**. The scores ranged from as little as 3 to the maximum 24. Of the 49 subjects, 5 scored less than 12, but 11 scored the full 24. 85% (42) of the subjects scored 17 or more out of the possible 24 and 77% of the subjects (38) scored 20 or more out of 24. The average times for each subject over all 24 trials ranged between approximately 30 and 90 seconds. It is interesting to note that over the high scores there is a wide range of average times whereas over the low scores there are relatively more high average times, suggesting that those people with low scores actually spent more time trying to find the correct solution.

The data in Figures 8 and 9 is ordered by the *overall number of successes by variant* in order to highlight any difference between performance over a particular diagram and performance overall. For each variant of each diagram, Figure 8 displays the total number of successes. The lowest number of successes occurs when Contour Jaggedness, Zone Area Inequality and Edge Closeness are all *high*, but the highest number is not when they are all *low*, although the difference between the successes for **lll** and both **lhl** (Zone Area Inequality high) and **hll** (Contour Jaggedness high) is small. Note that when Edge Closeness is the only attribute that is *high* (**llh**) the number of successes in the most complex diagram Euler.4.9 is far less than when both Contour Jaggedness and Zone area are *high* (**hhl**), giving a preliminary indication of the importance of edge closeness as the diagrams become more complex.

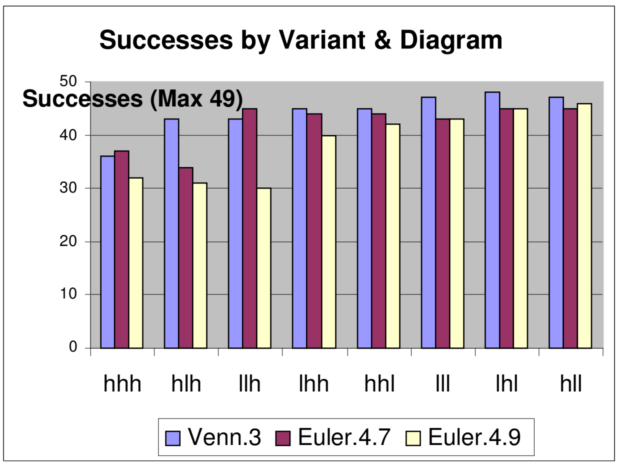

Figure 8

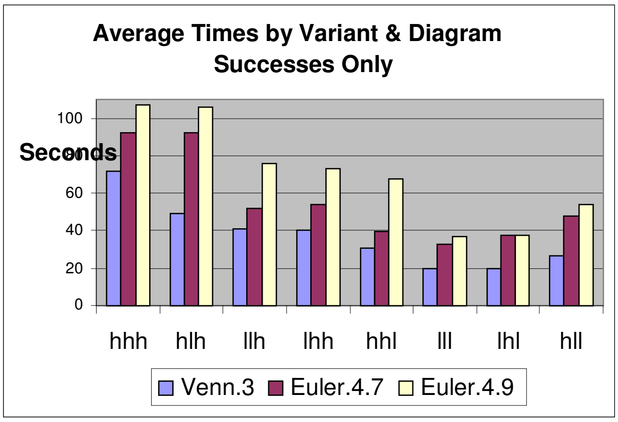

Figure 9

Figures 9 displays the average times, for each variant of each of the 3 diagrams, Venn.3, Euler.4.7 and Euler.4.9 ordered by the number of successes as in Figure 8. As in Figure 8 there is an indication, as expected that the more complex the diagram the more difficult it is to discern which zones are present and which are not. The shortest times are actually associated with the variant that is *low* for all three attributes (**lll**).

Statistical Analysis

The investigation was carried out with a randomised complete block design. Each of the 24 different diagrams was presented to each subject allowing a within subject design. The data is considered with respect to i) success or failure and ii) the time taken for each task. Tables 1 – 3 show which factors and interactions were significant with their *p*-values (most are <0.01). The *p*-value is the probability that the null hypothesis: the variation is random and that the factor has had no effect, is rejected when it is true.

Success or failure

Success is modelled as 1 and failure as 0. Since the dependent variable is discrete with only two possible values, the logistic regression model is used. All factors were taken into account including the session and the possible interactions between the diagrams and individual criteria and between the criteria themselves. The outcome of statistical analysis is shown in Table 1.

***p*-value**	**Factors**
<0.001	Subject, Diagram, Jagged, Edge
0.002	Session
0.004	Jagged.Zone
0.028	Jagged.Edge

Table 1 Success or Failure

Time taken

The distribution of the data over the time is slightly skewed, so the analysis is carried out over *ln(Time)* allowing a normal distribution. An analysis of variance performed over *ln(Time*) for all of the data returned similar significant effects to that over success or failure, as shown in Table 2.

***p*-value**	**Factors**
<0.001	Diagram, Jagged, Edge, Diagram.Jagged.Edge, Diagram.Unmooth.Zone
0.003	Zone
0.004	Diagram.Zone.Edge

Table 2 Analysis of Variance –Time (All data)

However, the times in Table 2 include those for incorrect solutions. Although this data is not inconsequential, there could be many reasons for incorrect solutions, apart from having difficulty in understanding the diagram. For example, the subject may simply misread the check box label or be concerned that the trial is taking too long. Hence an analysis was also conducted over the correct responses. An analysis of variance over the time taken as *ln(Time)* for correct solutions returns the effects shown in Table 3.

***p*-value**	**Factors**
<0.001	Diagram, Subject, Jagged, Edge, Zone, Diagram.Jagged.Edge, Diagram.Jagged.Zone
0.010	Diagram.Zone.Edge
0.028	Diagram.Edge

Table 3 Analysis of Variance – Time (Correct data)

The means over correct responses and *ln(Time)* for Degree, Year of Study and Gender do not indicate significant differences, but, of the 49 subjects, only 6 were studying joint honours degrees, only 6 were female and the number of subjects over the Years 1, 2, and 3 was 31, 17 and 1 respectively, so there is a lack of evidence with regard to these parameters.

3.2: Our interpretation of the results and discussion

A primary concern when evaluating subject responses, especially with respect to understanding is that the subject should understand the tasks they have been set. A score as low as 3 out of 24 suggests that the subject may not have understood the object of the tasks. However, the predominance (77%) of scores of 20+ is an indicator that for the most part the subjects did understand the nature of the tasks they were set.

The data by variant and by diagram (figures 8 and 9) serve to indicate that the diagrams vary in complexity of understanding as expected, namely, in ascending order of difficulty: Venn 3, Euler 4.7 and Euler 4.9. Since the nature of the task is such that the whole diagram must be inspected in order to find the solution, the ordering simply confirms expectation that as the number of contours increase and the number of zones increases, identifying the zones becomes harder.

The indication of an effect by Diagram, seen in figures 8 and 9 is also confirmed in tables 1, 2 and 3. From Table 1 it appears that Contour Jaggedness and Edge Closeness are more important than Zone Area Inequality, but as the times are taken into account first over all data and then over the correct solutions the importance of Zone Area Inequality becomes apparent. From all three statistical tests there is strong evidence to suggest that all three factors under consideration are important both as independent factors and as interactions with Diagram. Note that the interaction between Diagram, Zone Area Inequality and Edge Closeness is more apparent in the data over the correct solutions.

The interaction between Diagram, Contour Jaggedness and Edge Closeness (Tables 1 and 2) is also evident in Figure 8 which shows that when Contour Jaggedness and Edge Closeness are both *low* the number of successes for each diagram is almost the same; however, when both are *high*, i) the number of successes decreases and ii) the time taken for success increases, rapidly in proportion to the increase in complexity of the diagram.

The evidence here strongly suggests that all three of the chosen criteria affect the understanding of Euler diagrams, most particularly Contour Jaggedness and Edge Closeness. Closer inspection of the differences between the means for *ln(Time)* for correct solutions (success) allows an ordering on these criteria (ascending): Zone Area, Contour Jaggedness, Edge Closeness. However, given the evidence to suggest that interactions become more pronounced as the diagrams become more complex, it would not be sensible to predict a weighting between these criteria until further investigations have been carried out.

4: SUMMARY

This work is a preliminary step into using empirical evidence to support decisions concerning the metrics that mandate automated layout of Euler diagrams. Our investigation shows there is strong evidence to support the three chosen factors as important with regard to diagram layout.

It appears that the interactions between criteria become more pronounced as the diagrams become more complex, and in the light of this, further investigations could be conducted. By reducing the number of tasks and increasing the complexity of the diagrams it may be possible to qualify by degree the relationships between the various criteria of diagram layout and specify more precisely which interactions are the most important.

There is also a need for further work to expand the criteria investigated as other factors such as contour size and line intersection angle could affect the understanding of Euler diagrams. Another possible area of investigation relates to the notion that some Euler diagrams cannot be drawn without triple points, contours sharing line segments or contours taking figure of eight shapes, and it would be useful to discover the implications of such features on user comprehension.

Other important future work is in looking at the effectiveness of Euler diagrams in the context of application areas. This could include investigations examining how users interact with Euler diagrams when attempting to complete real world tasks. Many of the application areas rely on graph enhanced Euler diagrams, and so it would be useful to initiate investigations into the comprehension of these structures.

REFERENCES

1. John Bovey, Florence Benoy, and Peter Rodgers. Using Games to Investigate Movement for Graph Comprehension. Proc. Advanced Visual Interfaces: AVI 2004. pages 71-79. ACM, May 2004.
2. John Bovey, Peter Rodgers, and Florence Benoy. Movement as an Aid to Understanding Graphs. In Seventh International Conference on Information Visualization (IV03), pages 472-478. IEEE, 2003.
3. R. De Chiara, U Erra and V. Scarano. VENNFS: A Venn-Diagram File Manager. Proc. IEEE Information Visualization (IV03). pages 120-126. 2003.
4. S. Chow and F. Ruskey. Drawing Area-Proportional Venn and Euler Diagrams. Proc. GD2003. LNCS 2912. Springer Verlag.
5. J. Flower and J. Howse. Generating Euler Diagrams, Proc. Diagrams 2002, Springer Verlag, 61-75.
6. Jean Flower, Peter Rodgers, and Paul Mutton. Layout Metrics for Euler Diagrams. In Seventh International Conference on Information Visualization (IV03), pages 272-280. IEEE, 2003.
7. J. Gil, J. Howse and S. Kent. Towards a Formalization of Constraint Diagrams, Proceedings of Human-Centric Computing (HCC 2001) Stresa, Italy, IEEE Computer Society Press, 72-79. 2001.
8. David Harel. On Visual Formalisms. Communications of the ACM. 31(5). pp. 514-530. 1988.
9. Higraph web page: http://db.uwaterloo.ca/~gweddell/higraph/higraph.html.
10. P. Mutton, P. Rodgers and J. Flower. Drawing Graphs in Euler Diagrams. Proc. Diagrams 2004. LNAI 2980. Springer Verlag. 66-81.
11. H.C. Purchase, R.F. Cohen and M. James. Validating Graph Drawing Aesthetics. *Validating Graph Drawing Aesthetics*. GD95, LNCS 1027. Springer-Verlag, pages 435-446. 1995
12. A. Verroust and M.-L. Viaud. Ensuring the Drawability of Extended Euler Diagrams for up to 8 Sets. Proc. Diagrams 2004. LNAI 2980. Springer Verlag. 128-141.
13. C. Ware, H.C. Purchase, L. Colpoys and M. McGill. Cognitive measurements of graph aesthetics. Information Visualization, 1(2), pages 103-110, 2000.