Kent Academic Repository

Full text document (pdf)

Citation for published version

Benoy, Florence and Rodgers, Peter (2004) A Study into the Comprehension of Euler Diagrams. University of Kent.

DOI

Link to record in KAR

https://kar.kent.ac.uk/14125/

Document Version

UNSPECIFIED

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.
Users are advised to check **http://kar.kent.ac.uk** for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:
researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at **http://kar.kent.ac.uk/contact.html**

Computer Science at Kent

A Study into the Comprehension of Euler Diagrams

Florence Benoy and Peter Rodgers

Technical Report No. 14-04
July 2004

Copyright © 2004 University of Kent at Canterbury

Published by the Computing Laboratory,
University of Kent, Canterbury, Kent CT2 7NF, UK

A Study into the Comprehension of Euler Diagrams

Florence Benoy, Peter Rodgers

Computing Laboratory, University of Kent
Canterbury, Kent, CT2 7NF
`{P.M.Benoy, P.J.Rodgers}@kent.ac.uk`
`http://www.cs.kent.ac.uk/`

Abstract. We describe an empirical investigation into some of the layout criteria that may facilitate the comprehension of Euler diagrams. The three criteria under investigation were: smoothness, zone area equality and edge closeness. Subjects were asked to interpret diagrams that had *good* and *bad* levels of each criterion. The goal of this research is an initial step into using empirical evidence to support decisions concerning the metrics used for automated layout of Euler diagrams. 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 importance for these criteria.

1 Introduction

Euler diagrams are becoming a widely used technique in visualization systems. They have the ability to represent 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. The automatic layout of Euler diagrams has recently been investigated utilizing a multicriteria optimizing system [7]. However, the justification for the metrics that were used 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.

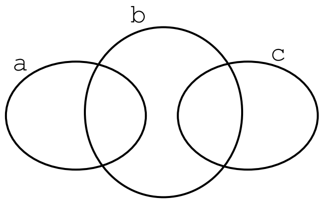

Fig. 1. An Euler diagram with countours: **a**, **b** and **c**; and zones: **a b c ab bc**

An 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.

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.

Euler diagrams are becoming of increasing use as a visualization technique. Recent application areas include: constraint diagrams [8], software modelling [11], the visualization of networks [9], database visualization [5] and file system organization [3]. In many of these applications the zones contain objects or graphs. When graphs are superimposed upon Euler diagrams, they can be seen as extended graph structures such as higraphs or hypergraphs, and so the layout of Euler diagrams enhanced with graphs can be applied to laying out these extended graph structures.

A challenging aspect of Euler diagram layout is embedding the diagram. Generating an embedding is not a fully solved problem. Flower and Howse [6] implement a mechanism for embedding Euler diagrams under strong wellformedness conditions. 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 [7]. This work was extended to Euler diagrams enhanced with graphs [10].

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 crossing are known to have an inhibiting effect on comprehension and consequently most algorithms will 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 [13,14]. These initial studies focused on graph theoretic questions concerning shortest paths in randomly drawn graphs. In the first instance the authors were able to demonstrate that, under the conditions of the experiment, reducing edge crossings and edge bends was significant in aiding understanding. In the context of social networks, where a primary purpose is facilitating the detection of paths (connections), the preferred layouts seem to favour symmetry and straight lines above the minimisation of line crossings [15].

The research question we address is to validate three criteria for improving the comprehension of Euler diagrams and, if possible, to infer an ordering 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 [7]. Several of these metrics were amalgamated into three criteria that were visually discernible as *good* or *bad*, namely: Smoothness, Zone Area Equality and Edge Closeness. The objective was to confirm, or otherwise, that these criteria play a significant part in facilitating the comprehension of Euler diagrams. The outcome of the investigation suggests that all three of these criteria, particularly Smoothness 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.

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.

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 diagrams had *good* or *bad* levels in all combinations of three criteria.

2.1 The Criteria

The criteria we chose to investigate were based on our experience with automatically laying out Euler diagrams. As the number of diagrams that need to be shown to subjects increases exponentially with the number of criteria, we limited the study to three criteria. We chose the three that seemed most likely to facilitate comprehension.

Figure 2 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 criteria we chose to use were: *Smoothness*, *Zone Area Equality* and *Edge Closeness*. When describing the diagram variants, the convention we use is to take the criteria in the order: Smoothness, Zone Area Equality and Edge Closeness, so that **bbb** is *bad* for all the criteria, whereas, **bgb** is *bad* for Smoothness and Edge Closeness, but *good* for the middle criterion, Zone Area Equality.

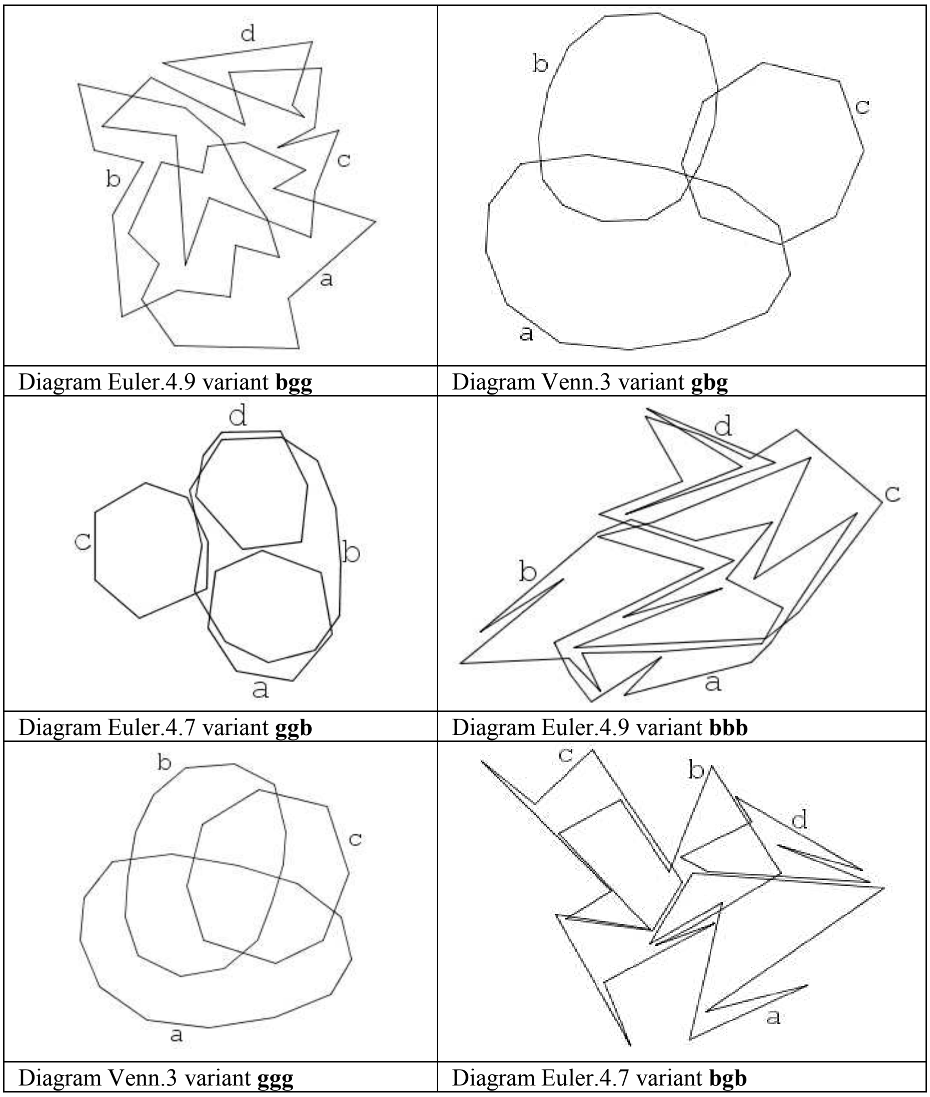

Diagram Euler.4.9 variant **bgg**

Diagram Venn.3 variant **gbg**

Diagram Euler.4.7 variant **ggb**

Diagram Euler.4.9 variant **bbb**

Diagram Venn.3 variant **ggg**

Diagram Euler.4.7 variant **bgb**

Fig. 2. Some example diagrams

The criteria in detail are:

- Smoothness relates to the continuousness of the contour lines, smooth lines, were rated *good* and jagged lines, *bad*. The diagram on the top left of Figure 2 has a *bad* smoothness rating, but is rated *good* for the other two criteria.
- Zone Area Equality relates to the relative sizes of the zone areas. An uneven distribution, with some zones very large and some very small is rated *bad*, an even distribution with all zones closer in size is g*ood*. The diagram on the top right of Figure 2 has a *bad* Zone Area Equality rating, but is rated *good* in the other two criteria.

- Edge Closeness relates to the closeness of lines from different contours. Diagrams with lines close together for large sections are rated *bad*, diagrams with lines always diverging are rated *good*. The diagram on the middle left of Figure 2 has a *bad* Edge Closeness rating, but is rated *good* 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 [7] 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 *good* measure to a *bad* one. The adjustments were also guided by re-measurement of the diagram to ensure that the visual change was reflected by an appropriate change in the metrics. 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 *good* and *bad* measures of zone area equality were uniformly wider than for the other two criteria.

2.3 Software

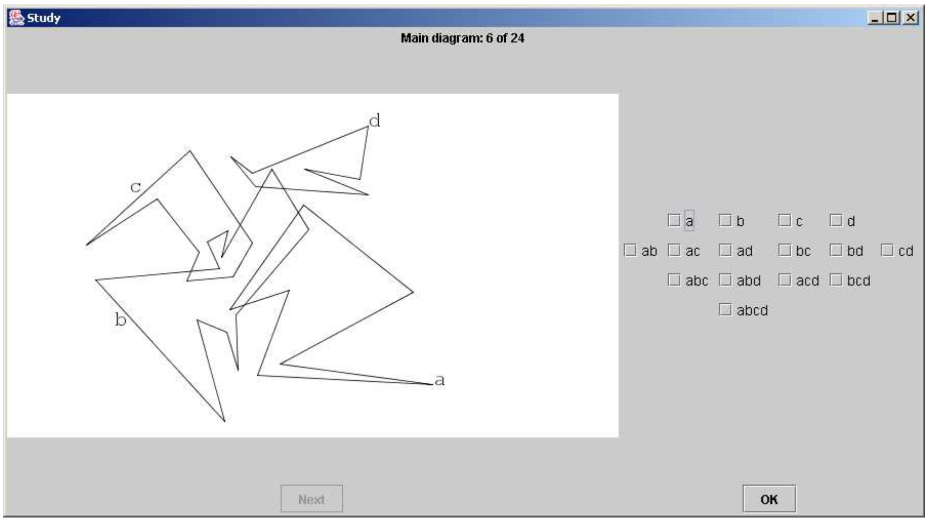

Fig. 3. This is a screenshot of the software with diagram Euler.4.7, variant **bbg**. The zones in this diagram are: **a b c d ab bc bd**

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 3 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.

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: Smoothness, Zone Area Equality and Edge Closeness. This gives a total of 24 main diagrams, some of which are shown in Figure 2. 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 demonstration of the task. The subjects were also given 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 performed 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 had become too familiar with the study investigations, as the majority of them had also been used for 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 variations of one diagram that the results of the pilot indicated were the most difficult. Whilst we were 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 is 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 the domain; set theory, taught using similar diagrams, is part of the first year of study.

3 Results

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

3.1 Summary of the Data

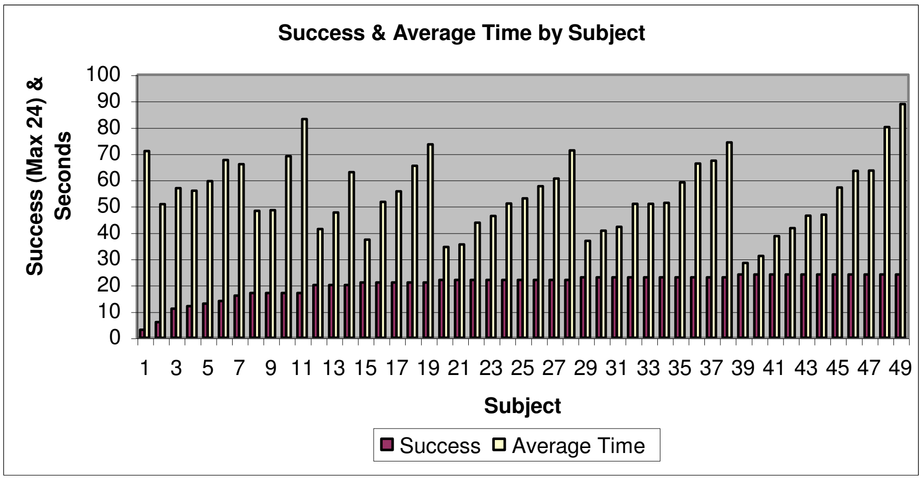

Fig. 4.

There were 49 subjects and figure 4 shows the very wide range of scores from as little as 3 to the maximum 24. 5 subjects scored less than 12, but 11 subjects scored the full 24. Figure 4 also displays the average time spent on each of the 24 trials by each subject (ordered by successes and then by average time). 85% (42) of the subjects scored 17 or more and 77% of the subjects (38) scored 20 or more. The average times for each subject over all of the 24 trials ranged between approximately 30 and 90 seconds. 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.

For each variant, Figure 5 displays the number of successes, and the average time and standard deviation over i) all of the data and ii) just the successes. The lowest number of successes occurs when Smoothness, Zone Area Equality and Edge Closeness are all *bad*, but the highest number is not when they are all *good*, although the difference between the successes for **ggg** and both **gbg** and **bgg** is small. Note that when Edge Closeness is the only attribute that is *bad* (**ggb**) the number of successes is less than for both **gbb** and **bbg**, both with two criteria measured as *bad*.

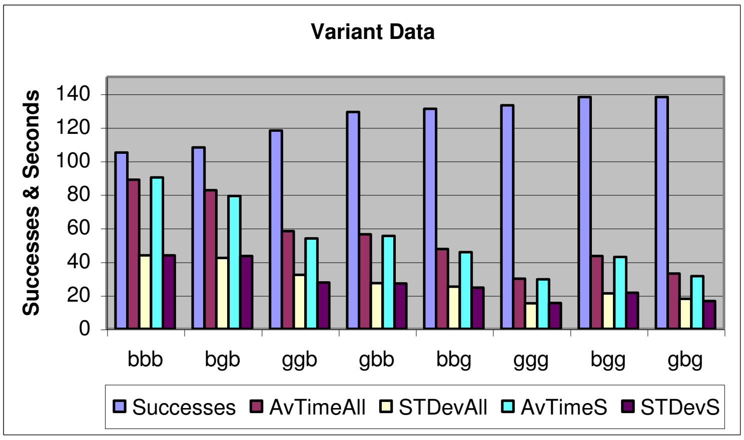

Fig. 5

Figure 5 shows a roughly inverse relationship between the number of successes and the average times for each variant. Note that for each variant the difference between the average time and standard deviation for all trials and the average time and standard deviation for just successes are for the most part in the region of 1 second which in terms of a human response that requires both cognition and a physical action is very small. The exception is for variant **ggb** (5 seconds difference for both average and standard deviation).

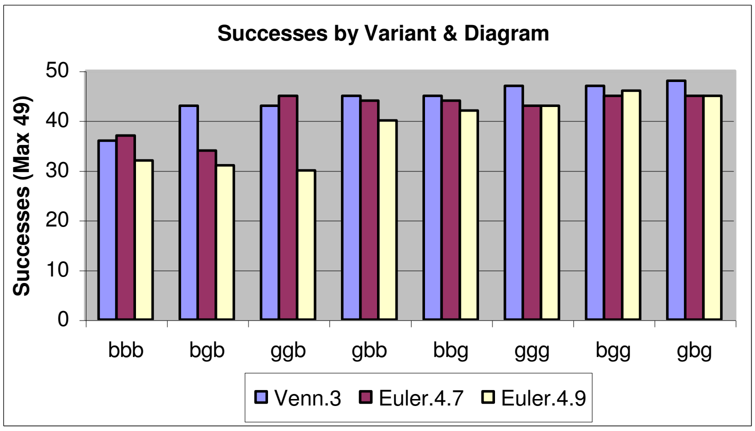

Fig. 6

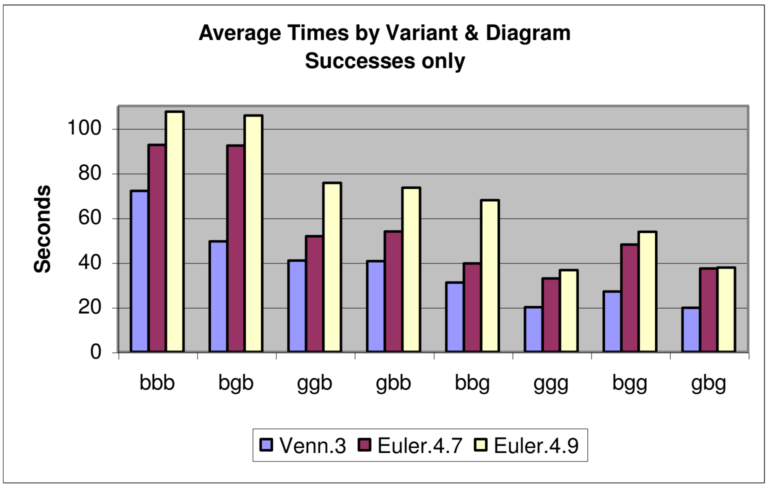

Fig. 7

Figures 6 and 7 display the successes and average times, respectively, for each variant of each of the 3 diagrams, Venn.3, Euler.4.7 and Euler.4.9. These figures are ordered by the overall number of successes by variant in order to highlight any difference between performance over a particular diagram and performance overall. The outcome for the individual diagrams is very close to that overall with two exceptions, both for Euler.4.9. In this case variants **ggb** and **bbb** have effectively swapped places as far as order is concerned although the actual difference in the number of scores is small 32 - 30 = 2. However inspection of the incorrect solutions for Euler.4.9, variant **ggb**, reveals that of the 18 incorrect solutions 8 were the same. This incorrect solution could be valid for an interpretation of the diagram that allows contours to touch. The occurrence highlights the fact that no mention was made explicitly in either the information sheet or in the preamble that such an occurrence could not legitimately occur in the presented Euler diagrams. The incorrect solutions for Euler.4.9, variant **bbb** have no similar incidence of the same solutions.

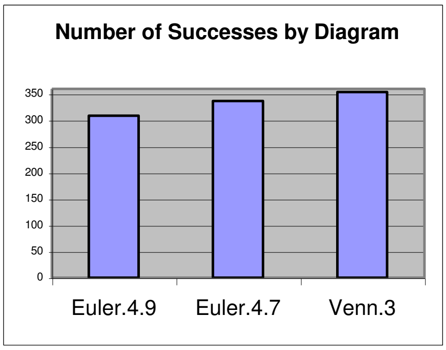

Fig. 8. Figure 8 displays the number of successes for each diagram

3.2 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: that 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.

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

Table 1

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, Smooth, Edge, Diagram.Smooth.Edge, Diagram.Smooth.Zone
0.003	Zone
0.004	Diagram.Zone.Edge

Table 2

However, the times above 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 following effects:

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

Table 3

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.

3.3 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 (Figure 4) suggests that the subject did not understand the object of the tasks. However, the predominance of scores of 20+ is an indicator that for the most part the subjects understood the nature of the tasks they were set.

The data by variant and by diagram (figures 5, 6 and 7) serve to indicate that the diagrams vary in complexity of understanding as expected and Figure 8 confirms the trends in the previous 3 figures that indicate an ordering of complexity over the diagrams. Since the nature of the task is such that the whole diagram must be inspected in order to find the solution, the ordering confirms 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 5 through 8 is confirmed in tables 1, 2 and 3. From Table 1 it appears that Smoothness and Edge Closeness are more important than Zone Area Equality, but as the times are taken into account first over all data and then over the correct solutions the importance of Zone Area Equality 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. It is interesting to note that the interaction between Diagram, Zone Area Equality and Edge Closeness is more apparent in the data over the correct solutions.

Smooth/Zone/Edge		**gbg**	**bgb**	**ggb**	**bbg**
Overall	Successes	138	108	118	131
	Time	31	79	53	45
Venn.3	Successes	48	43	43	45
	Time	20	50	41	31
Euler.4.7	Successes	45	34	45	44
	Time	37	92	51	40
Euler.4.9	Successes	45	31	30	42
	Time	38	105	75	68

Table 4

Closer inspection of Figures 5, 6 and 7 shows that when Smoothness and Edge Closeness are both *good*, relative to the successes over all diagrams, the number of successes for each diagram is almost the same, but the average time taken increases with the complexity of the diagram; further, when both are *bad*, the number of successes decreases and the average time increases with the complexity of the diagram. A similar, though not so pronounced trend can be seen with Smoothness and Zone Area. The successes and times are shown in Table 4 for comparison. This may indicate that as diagrams increase in complexity, the interactions between factors come into play.

The evidence here strongly suggests that all three of the chosen criteria affect the understanding of Euler diagrams, most particularly Smoothness 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, Smoothness, 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.

An important area of 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.

Acknowledgements

We thank Judith Masthoff, Jean Flower and John Howse at the University of Brighton, as well as Martin Ridout and Jelena Havelka at the University of Kent for their useful discussions about the work in this paper.

References

1. John Bovey, Florence Benoy, and Peter Rodgers. Using Games to Investigate Movement for Graph Comprehension. In Advanced Visual Interfaces: AVI 2004. ACM, May 2004. To Appear.
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). pp. 120-126. 2003.
4. S. Chow and F. Ruskey. Drawing Area-Proportional Venn and Euler Diagrams. Proc. GD2003. LNCS 2912. Springer Verlag.
5. M.P. Consens and A.O. Mendelzon. Hy+: A Hygraph-based Query and Visualization System. In Proc. ACM SIGMOD Intl. Conf. on Management of Data, pp. 511-516, 1993.
6. J. Flower and J. Howse. Generating Euler Diagrams, Proc. Diagrams 2002, Springer Verlag, 61-75.
7. 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.

8. 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.
9. GXL web page: http://www.gupro.de/GXL/examples/ hypergraphNav.html.
10. P. Mutton, P. Rodgers and J. Flower. Drawing Graphs in Euler Diagrams. Proc. Diagrams 2004. LNAI 2980. Springer Verlag. 66-81.
11. M.-A. D. Storey and H. Mueller. Manipulating and documenting software structures using SHriMP views. In Int. Conf. in Software Maintenance, pp. 275-285. IEEE. 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. 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
14. C.Ware, H.C.Purchase, L.Colpoys and M.McGill. Cognitive measurements of graph aesthetics. Information Visualization, 1(2), pages 103-110, 2002.
15. U.Brandes and D.Wagner. Analyis and Visualisation of Social Networks Proceedings of the 2nd International Workshop on Experimental and Efficient Algorithms (WEA'03). LNCS 2647, pages 261-266. Springer-Verlag, 2003.