Evaluating the Comprehension of Euler Diagrams

Florence Benoy
University of Kent

Peter Rodgers
University of Kent

ABSTRACT

We describe an empirical investigation into layout criteria that can help with the comprehension of Euler diagrams. Our work is intended to inform automatic Euler diagram layout research by confirming the importance of various Euler diagram aesthetic criteria. The three criteria under investigation were: smoothness, zone area equality 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 importance for the criteria.

1. INTRODUCTION

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. Studies that seek to validate (or otherwise) commonly used criteria by empirical investigation have been pioneered by Purchase [5].

The automatic layout of Euler diagrams has only recently been investigated. A multicriteria optimizing system was developed [4], 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 [3] implement a mechanism for embedding Euler diagrams under strong wellformedness conditions. As the wellformedness conditions are relaxed, more diagrams can be drawn [2,6]. These embeddings are not typically aesthetically pleasing, and hence work has been performed in improving the layout of embedded diagrams [4]. 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 with no given interpretation. 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 [4]. Several of these metrics were amalgamated into three criteria: Smoothness, Zone Area Equality and Edge Closeness. 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. 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. 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. 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, we kept the design as simple as

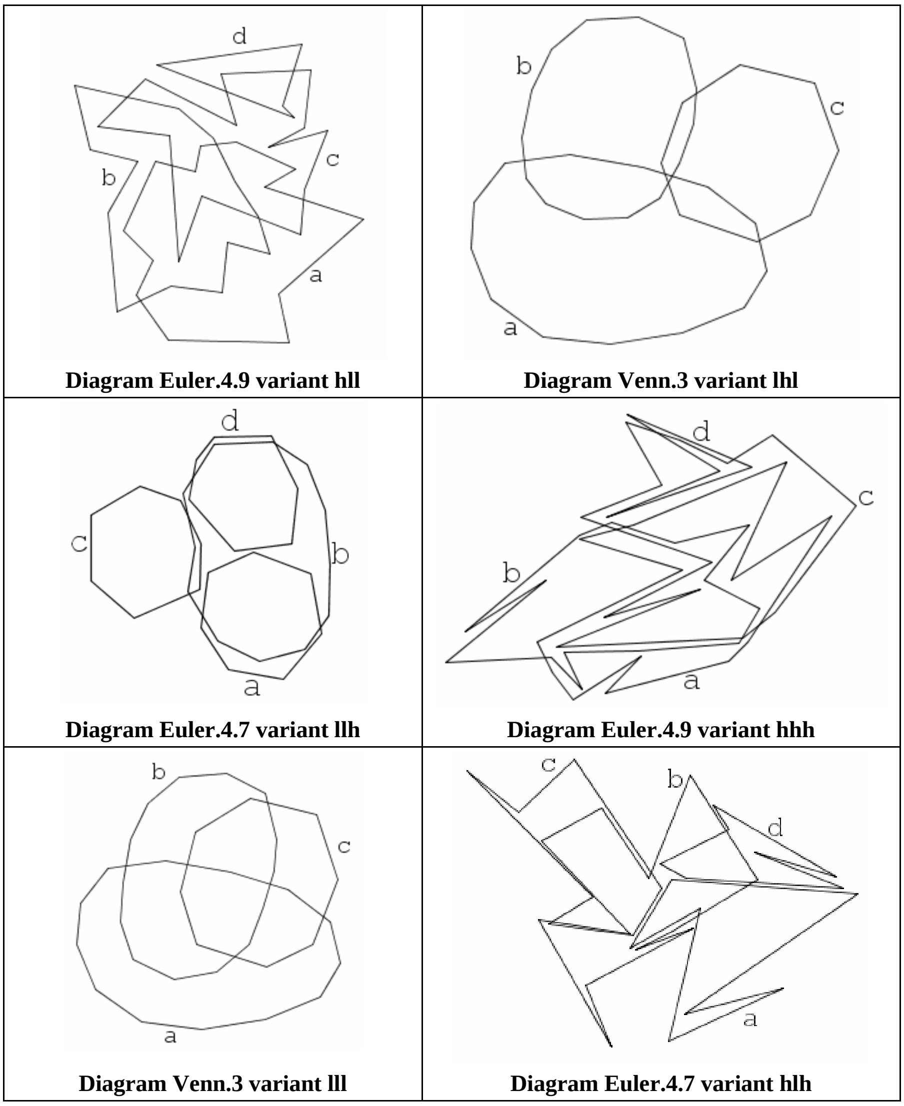

Figure 1. Some example diagrams.

possible whilst still being useful and limited the study to just three criteria. We chose the three that seemed most likely to facilitate comprehension, based on 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 of our subjects given the task. The complexity range of the diagrams shown in the main study was informed by pilot studies. For example, we withdrew one of our initial set of diagrams altogether, as too complex. Section 2.4. Figure 1 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: *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 **hhh** is *high* for all the criteria, whereas, **hlh** is *high* for Smoothness and Edge Closeness, but *low* for the middle criterion, Zone Area Equality.

The criteria in detail are:

- Smoothness relates to the continuousness of the contour lines. Smooth lines have a relatively *low* measurement and jagged lines a *high* one. The diagram on the top left of Figure 1 has a *high* smoothness measurement, but is rated *low* 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 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 1 has a *high* Zone Area Equality 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 1 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 [4] 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.

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. The subject checks 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: Smoothness, Zone Area Equality and Edge Closeness. This gives a total of 24 main diagrams, some of which are shown in Figure 1. The subjects were given one of 24 randomized sequences of variants (coincidentally the number of sequences is the same as the number of variants). 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 and demonstration of the task with the opportunity to ask questions. Eight training diagrams, each with feedback on performance and the correct solution preceded the main study. At the end of the session the subjects were given their overall result on-line with their performance detail and the solutions. Finally they filled in a questionnaire and read a short debriefing document explaining the nature of the 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 study 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.

Prior to the main study we carried out two pilot studies with 4 diagrams to check our methodology. From these it appeared that one diagram was more difficult than the others, hence we removed the relevant 8 variants, reducing the number of main test diagrams from 32 to 24.

The subjects for our main study were undergraduate computing students because they were accessible and have some knowledge of set theory, taught using similar diagrams in their first year. Hence, they would require only an introduction to the problem specification and environment. Further these students could well be representative of the population from which practitioners who either use or generate 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, i) a brief summary of the data, ii) the statistical results and iii) an interpretation and discussion of the outcomes.

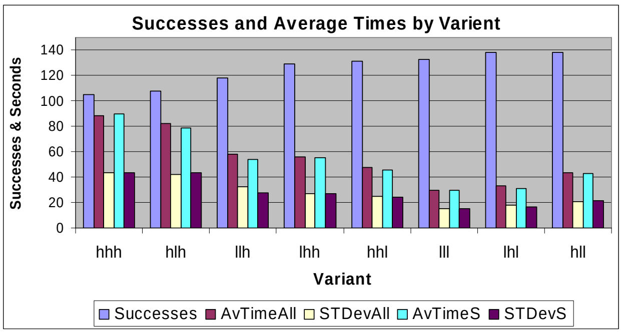

Figure 2. This shows a roughly inverse relationship between the number of successes and the average times for each variant.

3.1. Summary of the Data

To recap, the expectation for all three metrics is that a low rating is likely to facilitate comprehension, so, the **lll** variant is a smooth, equal zoned diagram with edges well apart For the purpose of this discussion a subject's ***score*** is the number of tasks completed successfully. There were 49 subjects who had a 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. For each variant, Figure 2 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 *high*, but the highest number is not when they are all *low*, although the difference between the successes for **lll** and both **lhl** and **hll** is small.

Note that when Edge Closeness is the only attribute that is *high* (**llh**) the number of successes is less than for both **lhh** and **hhl**, both with two criteria measured as *high*. Figure 2 also shows a roughly inverse relationship between the number of successes and average times for each variant

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. Table 1 shows which factors 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 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 (there were two) 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, Smooth, Edge
0.002	Session
0.004	Smooth.Zone
0.028	Smooth.Edge

Table 1. Success or Failure.

An analysis of variance over the time taken as *ln(Time)* for correct solutions returns the effects shown in Table 2.

***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 2. Analysis of Variance – Time (Correct data).

3.2. Our interpretation of the results and discussion

The data by variant (Figure 2) serves 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 confirms that as both the number of contours and zones increases, identifying the zones becomes harder.

From tables 1 and 2 it appears that Smoothness and Edge Closeness are more important than Zone Area Equality, but as time is taken into account 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.

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.

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. S. Chow and F. Ruskey. Drawing Area-Proportional Venn and Euler Diagrams. Proc. GD2003. LNCS 2912. Springer Verlag.
3. J. Flower and J. Howse. Generating Euler Diagrams, Proc. Diagrams 2002, Springer Verlag, 61-75.
4. 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.
5. 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
6. 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.