The Impact of Shape on the Perception of Euler Diagrams

Andrew Blake[1], Gem Stapleton[1], Peter Rodgers[2], Liz Cheek[1], and John Howse[1]

[1] University of Brighton, UK
{a.l.blake,g.e.stapleton,liz.cheek,john.howse}@brighton.ac.uk
[2] University of Kent, UK
{p.j.rodgers}@kent.ac.uk

Abstract. Euler diagrams are often used for visualizing data collected into sets. However, there is a significant lack of guidance regarding graphical choices for Euler diagram layout. To address this deficiency, this paper asks the question 'does the shape of a closed curve affect a user's comprehension of an Euler diagram?' By empirical study, we establish that curve shape does indeed impact on understandability. Our analysis of performance data indicates that circles perform best, followed by squares, with ellipses and rectangles jointly performing worst. We conclude that, where possible, circles should be used to draw effective Euler diagrams. Further, the ability to discriminate curves from zones and the symmetry of the curve shapes is argued to be important. We utilize perceptual theory to explain these results. As a consequence of this research, improved diagram layout decisions can be made for Euler diagrams whether they are manually or automatically drawn.

1 Introduction

An Euler diagram represents sets using graphical elements called closed curves. The interior of each curve represents elements that are in the set [20,22]. Fig. 1 shows a staff hierarchy within an academic institution. The (curve labelled) 'Managers' intersects with 'Academics' meaning that there are some managers who are academics. 'Researchers' is a subset of 'Academics' meaning that all researchers are academics. 'Researchers' is disjoint from 'Managers' so there are no researchers who are managers. Euler diagrams are often regarded a natural [22] and effective [21] way to depict sets. Their numerous application areas include the natural sciences [9], art and architecture [2], education [13], criminology [10], computer file organisation [6] and classification systems [26]. This provides clear and strong motivation for the need to better understand how the choices made when laying out (drawing) Euler diagrams impact on user comprehension. Providing such an understanding will improve the effectiveness of these diagrams as a mode of information visualization with, potentially, wide ranging benefits.

Reflecting their widespread use, recent times have seen a variety of methods derived to automatically produce Euler diagrams. These methods make varying choices of the topological and graphical properties to be possessed by the diagrams that they produce. For example, Wilkinson's method uses circles, motivated by the fact that 72 Euler diagrams used in articles appearing in Science, Nature and online affiliated journals, 90% use circles [28]. Wilkinson is not alone in choosing to use circles, with other automated drawing methods doing the same, such as [24]. However, Micallef and Rodgers [16] prefer the use of ellipses while Riche and Dwyer [19] prefer the use of rectangles, albeit in a stylized form. Other methods make no preference towards any particular geometric shapes at all [23] but some of them, including [25], aim to minimize or avoid certain topological properties, such as the use of non-simple curves. There is a need to derive more informed automated layout methods that take proper account of user comprehension when drawing Euler diagrams. At present, many of the graphical choices are based on assumptions about what yields an effective diagram.

As there are a range of automated Euler diagram drawing methods, using a variety of shapes, it seems important to establish the relative performance of users with different shapes. Providing such an understanding will allow those devising drawing methods to prioritize efforts towards the use more effective shapes. The contribution of this paper is guidance on shape choice by conducting an empirical study. The next section provides a discussion on related work on Euler diagram drawing choices, covering the state-of-the-art in terms of layout guides. Section 3 presents the experiment design, followed by section 4 where experiment execution is described. The data are analyzed in section 5. We interpret the significant results with reference to perceptual theory in the section 6. Threats to the validity of the experiment are described in section 7, after which we conclude in section 8. All of the diagrams used in our study, and the data collected, are available from `http://www.cem.brighton.ac.uk/staff/alb14/experimental_resources/shape/shape.html`

2 Euler Diagram Background

Given a data set to be visualized, there are numerous choices of Euler diagram that can be drawn. We categorize these choices into three types: descriptional (the abstract syntax level), topological and graphical (both at the concrete syntax level). To illustrate, Figs 2 and 3 show four Euler diagrams that each represent the same information yet vary the choices made. When starting with the information that we wish to visualize using an Euler diagram, the first choice that

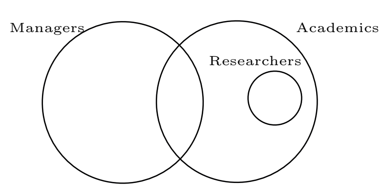

Fig. 1. Staff roles.

must be made is descriptional, at the so-called abstract syntax level. Descriptional choices determine the zones that must be present in the Euler diagram [24]. To illustrate, suppose we wish to represent the following syllogism:

1. All Students are People
2. All People are Mammals
3. (therefore) All Students are Mammals.

The Euler diagram on the left of Fig. 2 represents the above syllogism. The curve labelled 'S' represents students, 'P' represents people and 'M' represents mammals."All Students are People" is expressed by drawing 'S' inside 'P'. Similarly, "All People are Mammals" is expressed by drawing 'P' inside 'M'. From this diagram we can 'read off' "All Students are Mammals". This diagram can be described by an abstract syntax, which is a list of zone descriptions: $\emptyset$, M, MP, and MPS. For instance, $\emptyset$ describes the zone that is outside all three curves, and MP describes the zone that is inside 'P' and 'M' but outside 'S'. Other Euler diagrams can represent the same syllogism. They have zones for $\emptyset$, M, MP, and MPS as well as additional zones which are shaded to assert no elements are in the set they represent, an example of which is illustrated on the right of Fig. 2. The left diagram is considered, by Gurr [11], to be the most effective, precisely encapsulating the semantics of the syllogism. This diagram is said to be *well-matched* to its meaning as it has no additional zones. Gurr explains:

> "The transitive, irreflexive and asymmetric relation of set inclusion is expressed via the similarly transitive, irreflexive and asymmetric visual of proper spatial inclusion in the plane [11]."

To summarise, we define guide 1 for Euler diagram drawing:

Guide 1 (Well-matched) *Draw well-matched Euler diagrams (i.e. no extra zones).*

Being able to draw a well-matched diagram only solves part of the problem of ascertaining an effective layout. Fig. 3 illustrates two further well-matched diagrams. However, both diagrams exhibit topological properties, our second category of choice type, known to inhibit effectiveness. Diagram 1 shows the following topological properties. Curves 'P' and 'M' illustrate brushing points,

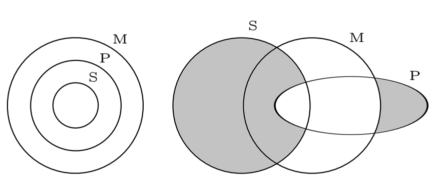

Fig. 2. Choices of abstract syntax.

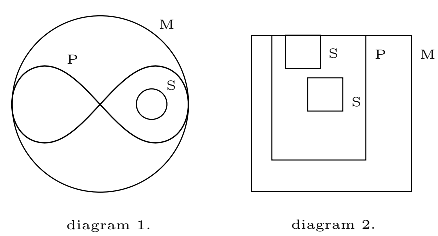

Fig. 3. Choices of topology.

defined as two or more curves which meet but do not cross at a point. Curves 'P' and 'M' give rise to a disconnected zone, defined as a zone that consists of one or more minimal regions. Curve 'P' is a non-simple (self-intersecting) curve. In diagram 2, curve 'P' and one curve labelled 'S' have concurrency, defined as two or more curve segments sharing the same line. Curves 'M', 'P' and one labelled 'S' illustrate a triple point, defined as three curves meeting at a point. The two curves labelled 'S' illustrate duplicated curve labels, where two or more curves represent the same set. Diagrams exhibiting none of these properties are *well-formed*. Rodgers et al.'s study [20] considers the impact of well-formed properties, summarised here as guide 2:

Guide 2 (Well-formed) *Draw well-formed Euler diagrams.*

Irrespective of laying out well-matched and well-formed Euler diagrams, there still exist numerous graphical choices to be made, our third category of choice type. Benoy and Rodgers [3], with their work on aesthetics, acknowledged the importance of making the correct graphical choices when drawing Euler diagrams. They conducted a study that focused on the jaggedness of curves, zone area equality and the closeness of one closed curve to another. To summarise their results, we define three further guides:

Guide 3 (Smooth curves) *Draw Euler diagrams with smooth curves.*

Guide 4 (Zone area equality) *Draw Euler diagrams with zone area equality.*

Guide 5 (Diverging lines) *Draw Euler diagrams with diverging lines.*

Benoy's and Rodgers' work, while valuable, remains limited. There are many other graphical choices that might be considered. Bertin [4] identifies both retinal and planar variables, which constitute a variety of graphical choices, to which we are known to be perceptually sensitive. With respect to planar variables, we established that the effect of an Euler diagram's orientation does not impact on users' comprehension [5], leading to:

Guide 6 (Orientation) *Draw Euler diagrams without regard to orientation.*

Retinal variables include shape and colour. With resect to shape, Figs 4, 5, 6 and 7 illustrate four equivalent diagrams each adhering to guides 1 to 6 but drawn using different shapes. None of the aforementioned automated layout methods have been shown to adhere to all six guides. The remainder of this paper describes our study undertaken to establish whether the shape of a closed curve – in particular squares, circles, ellipses and rectangles – affects users' comprehension.

3 Experimental Design

For the purposes of this study, congruent with previous studies [5,14,17,18,19], we view comprehension in terms of task performance: one diagram is more comprehensible than another diagram if users can interpret it, on average, more

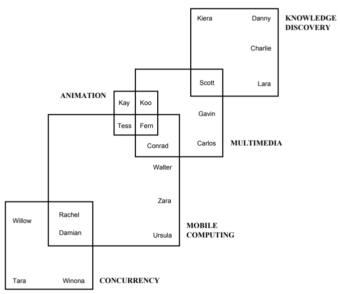

Fig. 4. Type 1 diagram: squares.

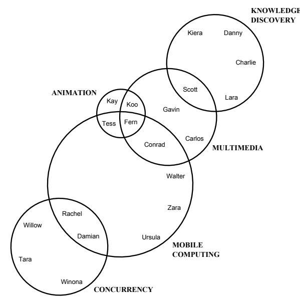

Fig. 5. Type 1 diagram: circles.

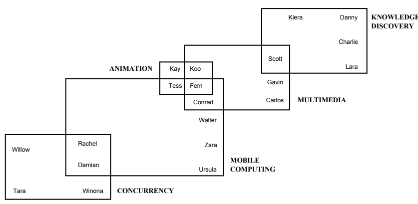

Fig. 6. Type 1 diagram: rectangles.

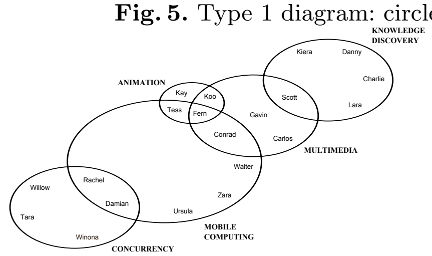

Fig. 7. Type 1 diagram: ellipses.

quickly and with fewer errors. We compare squares, circles, ellipses and rectangles. These four shapes were identified because they are pervasive in existing Euler diagram layout work and are widely used. We adopted a between group design consisting of four participant groups. All participants were asked questions of 18 Euler diagrams. If shape impacts on comprehension then we would expect to see, for some diagram, significant differences between time taken or errors accrued. From this point forward, all Euler diagrams in the paper and are scaled versions of those used in the study. Next, we discuss a series of factors considered during the design. All of the diagrams in the study were drawn sensitive to the six layout guides and the following drawing conventions, to ensure that different diagrams had consistent layout features:

1. all diagrams were drawn using 3 sized curves: small, medium and large,
2. the medium and large curves were scaled 200% and 300%, respectively, relative to the small curve,
3. rectangles and squares were drawn with their sides parallel to the x and y axes and, similarly, so were the major and minor axes of the ellipses,
4. each rectangle and ellipse was drawn adhering to the golden ratio,
5. all closed curves were drawn with a 2 pixel stroke width,
6. all diagrams were monochrome, drawn in an area of 810×765 pixels,
7. the curve labels were written using upper case letters in Times New Roman, 14 point size, font in bold,
8. data items were written using lowercase letters, except that the first letter was capitalised, and with Ariel 12 point size font,
9. each curve label was positioned closest to its corresponding curve, and

10. data items were evenly distributed within each zone.

We also required a variety of diagrams to be drawn. To this end, diagrams were drawn pertaining to the following three characteristic types:

1. Type 1: 5 curves (1 large, 3 medium and 1 small), 11 zones and 20 data items (Figs 4 to 7),
2. Type 2: 7 curves (2 large, 3 medium and 2 small), 15 zones and 30 data items (Fig. 8) and
3. Type 3: 9 curves (3 large, 3 medium and 3 small), 19 zones and 40 data items (Fig. 9).

In our study, and congruent with [20], a real-world scenario was employed as it was regarded pertinent to the reader. Further, abstract representations were considered a barrier the participants understanding. Therefore, information was visualized about fictional university modules and the students. It was anticipated that participants used in the study would be university students and therefore would have a reasonable understanding of this information. The module names were based on those commonly found in British undergraduate computing courses. Student names were first names only, a mixture of both male and female names, and reflected a variety of ethnicities. Following [20], three styles of question were specified, 'Who', 'Which' and 'How', that allowed us to elicit the following type of information:

1. Who is taking ANIMATION, MULTIMEDIA and MOBILE COMPUTING?
2. Which module is being taken by 5 students?
3. How many students are taking DATA STRUCTURES and HCI but not MARKETING?

The above questions were asked of Figs. 4, 8 and 9 respectively. All questions had a choice of five answers, a unique one of which was correct. The diagrams

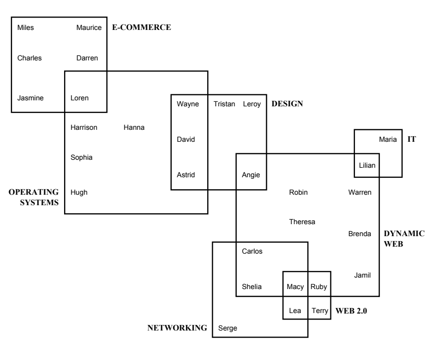

Fig. 8. Type 2.

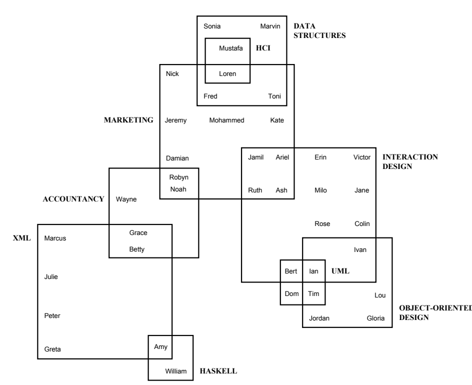

Fig. 9. Type 3.

were divided equally into the three characteristic types, each allocated two of each question style.

Initially, 18 diagrams were drawn using squares and data items were added to each diagram. The diagrams drawn for the other three shapes were equivalent in terms of their underlying description and topology as those drawn using squares. Given a curve in a 'squares' diagram, the corresponding curves in the other three diagrams had the same area. Careful attention was given to how the syntax was positioned: the placement of data items and curves labels was preserved, as much as possible, across equivalent diagrams.

For the collection of performance data, we used a software tool (called the research vehicle) to display the diagrams and questions, to gather answers and the time taken. The time taken to answer a question was determined from the instant a question was presented until the instant a participant had selected an answer to the question. Each time the participant answered a question, the research vehicle would ask them to indicate when they were ready to proceed to the next question, thus allowing a pause between questions. There was a maximum time limit of two minutes for each question to ensure that the experiment did not continue indefinitely. The 18 diagrams were presented in a random order.

To collect preferential data, participants were asked to rank four diagrams based on their shape. A scale of 1 to 4 was employed with 1 being the most preferred shape and 4 being the least preferred shape. If desired, shapes could be ranked equally. Participants were also required to explain the reasoning for their ranking. Initially, participants were asked to rank each diagram based on their aesthetic preference. They were then asked to rank each diagram based on how they perceived the ease of answering a 'Who', 'Which' and 'How' style question respectively.

4 Experiment Execution

Having conducted two pilot studies and implemented a number of minor adjustments, eighty participants were recruited for the main study. All 80 participants were randomly allocated to groups each with 20 participants. They were all students from the University of Brighton's School of Computing, Engineering and Mathematics and they spanned both undergraduate and postgraduate levels. The experiment was performed within a usability laboratory which affords a quiet environment free from noise and interruption. The same computer and monitor was used by each participant. All participants were alone during the experiment, in order to avoid distractions, with the exception of an experimental facilitator who was present throughout. All participants that took part in the study successfully completed the experiment. The experiment took approximately 1 hour per participant and they were given £6 for their contribution to the research. Next we discuss the four phases of the experiment.

The first phase of the experiment was initial training. All participants were asked whether they were familiar with Euler diagrams to which the response was no but some stated they had seen Venn diagrams. Consequently, all participants were treated as having no previous experience of Euler diagrams and given the same training. Training began by introducing participants to the notion of Euler diagrams and the types of questions to be asked. This was achieved using hard copy printouts of the diagrams, one for each style of question. Participants were given a few minutes to study the diagrams and questions, after which the experimental facilitator explained how to answer the questions. The second phase of the experiment provided participants with further training on the notion of Euler diagrams as well as how to use the research vehicle. Participants were presented with six questions, one at a time. If a question was answered incorrectly the facilitator went through the question with the participant. The third phase of the experiment is where we collected performance (time and error) data. Lastly, if desired, participants took a short break before entering the final, fourth phase of the study where we collected preferential data.

5 Statistical Analysis

The following analysis of the performance data is based on 80 participants each attempting to answer 18 questions: $80 \times 18 = 1440$ observations. The grand mean was 20.35 seconds and the mean times taken to answer questions by shape are: squares 19.48; circles 15.96; ellipses 21.96; and rectangles 24.01.

In order to establish if there is significant variation across shapes, we conducted an ANOVA. The results of the ANOVA are summarised in table 1, which uses log (base 10) of time to achieve normality. If $p \leq 0.05$ then we regarded the result as significant. The row for shape, with a p-value of 0.008, tells us that there are significant differences between the mean times of at least one pair of shapes. Hence, we can see that shape impacts user comprehension. The row for question, with a p-value less than 0.001, tells us that there are significant differences between the mean times of at least one pair of questions. This shows robustness, in that the questions are sufficiently varied, requiring different amounts of user effort to answer. The row for shape*question, with a p-value of 0.042, tells us that there are significant differences between the mean times for at least one question and an associated pair of shapes. This allows us to deduce that there is a question for which the mean times taken for the associated four diagrams are significantly different. Again, this reinforces the observation from row 1 that shape impacts user comprehension. Further, when removing timeouts from the data (i.e. when participants failed to provide an answer within the two minutes allowed), of which there were only two, the above results are unchanged.

Next, we performed a Tukey simultaneous test which compares pairs of shapes, to establish whether one mean is significantly greater than another in order to rank the shapes. P-values of less than 0.01 were regarded to be significant, given multiple comparisons being made on the same data. All comparison returned a p-value of less than 0.001 with the exception of the comparison between ellipses and rectangles ($p = 0.066$). Table 2 presents the rankings of shape. We see that circles allow participants to perform significantly faster than with squares which, in turn, are significantly faster than both ellipses and rectangles.

However, there was no significant difference in performance between ellipses and rectangles with respect to time taken.

Where there were significant differences between pairs of shapes, the magnitude of these differences is reflected in the following effect sizes, given in table 3. For example, the largest effect size tells us that 73% to 76% of participants were faster interpreting diagrams drawn with circles than the average person using rectangles. By contrast the smallest (significant) effect size is between squares and ellipses at 54% to 58%, which is still a substantial difference. Thus, we conclude that circles are the most effective shape, with respect to time taken, followed by squares and then, jointly, by ellipses and rectangles.

Regarding errors, of the 1440 observations there were a total of 62 errors (error rate: 4%). The errors were distributed across the shapes as follows: squares 6; circles 14; ellipses 16; and rectangles 26. We performed a chi-square goodness-of-fit test to establish whether shape had a significant impact on the distribution of errors. The test yielded a p-value of 0.003. However, we noted an outlier with respect to rectangles. A single participant yielded approximately one third of all the errors accrued under rectangles, thus potentially biasing the results. Moreover, one question accounted for 16 of the errors. Consequently, we treat this p-value with caution. Re-running the test with the outliers removed yields a p-value of 0.094, leading us not to reject the null hypothesis: shape does not impact on error rate. Taking both error rate and time into account we have, therefore, evidence that circles are significantly more effective than squares, ellipses and rectangles.

The following analysis of the preferential data concerning shape is based on 80 participants answering 4 questions. Recall, participants were asked to rank the shapes by aesthetic preference, and the perceived ease of answering the three styles of question. To analyze the preferential data, we performed four Friedman tests. The p-values are: aesthetic 0.013; who 0.006; how 0.049; and which 0.000. Hence there are significant differences between the shapes in all four cases. Consequently, four Wilcoxon signed ranked tests were performed to identify the significant differences between pairs of shapes. As with the pairwise comparisons above, p-values of less than 0.01 are regarded to be significant because multiple comparisons being made on the same data. The rankings are presented in table 4. To summarize, circles were ranked top in all cases except for the 'Which' case. For 'Which' style questions, participants perceived square

Source	DF	MS	F	P
shape	3	1.41310	4.22	0.008
question	17	1.08369	48.88	0.000
shape*question	51	0.03057	1.38	0.042
participant(shape)	76	0.33513	15.12	0.000
Error	1292	0.02217		
Total	1439			

Table 1. ANOVA for the log of time.

Shape	Number	Mean (log)	Ranking
Circle	360	1.161	A
Square	360	1.224	B
Ellipse	360	1.276	C
Rectangle	360	1.304	C

Table 2. Pairwise comparisons.

Shape	Circle	Square	Ellipse	Rectangle
Circle	–	58%-62%	66%-69%	73%-76%
Square	–	–	54%-58%	62%-66%

Table 3. Effect sizes.

Shape	Aesthetic	Who	How	Which
Circle	A	A	A	B
Square	B	B	A B	A
Ellipse	B	A B	B	C
Rectangle	A B	A B	A B	A

Table 4. Wilcoxon tests.

and rectangles to be preferable to circles and ellipses. Analysing the qualitative data gathered in phase 4, participants' reason for this preference was that they found it easier to count data items that were "listed" or in a "tabular" form, as it was presented for squares and rectangles, as opposed to data items that were considered to be "randomly" distributed within circles or ellipses; participants' reasons were not primarily based on the shapes of the curves.

To investigate this further, we performed another pairwise comparison restricted to just the time data for 'Which' questions. This revealed that circles were still significantly more effective than all other shapes, whereas rectangles were the least effective shape. There was no significant difference between ellipses and squares. Given that the study was not designed for this analysis, only having six 'Which' style questions asked of the diagrams, it would not be robust to accept this as a significant difference. Thus, a new study may be required to establish whether differences in performance exist when data items are "listed" or in "tabular" form in all shapes, not just for squares and rectangles.

6 Interpretation of Results

Our results suggest that, overall, circles are significantly more effective than the other shapes. They were significantly faster for accessing information and in doing so accrued an insignificant number of errors, comparatively. For all curve shapes, information is accessed from within a closed curve or from zones. With this in mind, we next refer to perceptual theory to provide insight into the manifestation of our results. We place emphasis upon Gestalt principles [15,27] regarded to be central to shape discrimination. We also consider similarity theory [7,12] as it identifies important constraints upon the speed of a visual search.

First, we specifically refer to the principle of good continuation and the time taken, during a visual search, due to any similarity between targets and distracters. The principle of good continuation states that shapes consisting of smooth curves are easier for the eye to follow than those shapes made up of 'hard' or jagged contours. With respect to curve smoothness, we have already seen guide 3 (smooth curves). Circles and ellipses both adhere to this guide. However, squares and rectangles do not because they have corners, meaning that the principle of good continuation is contravened.

Largely because they are smooth, both circles and ellipses give rise to zones that take rather different shapes to the curves; the boundaries of the zones are typically not smooth. In particular, the point at which one circle/ellipse intersects another manifests a sudden large change in good continuation: the two intersecting smooth curves are abruptly discontinued at the point at which they intersect. Large changes in good continuation are said to promote shape discriminability and, during the process of figure-ground segregation, we posit that intersections between circles/ellipses stand out to become salient. By contrast, the intersections in diagrams drawn with squares and rectangles are not salient: they look similar to the corners of the curves, as illustrated in Fig. 4, and are not therefore easily discriminable from the curves. These observations have implications for discriminating the zones from the curves. We posit that, due to the similarity between squares/rectangles and their (rectilinear) zones, the zones and curves are not discriminable. By contrast, the zones in diagrams drawn with circles and ellipses are discriminable.

Now, similarity theory [7] states that search time increases based on two criteria. The first criteria pertains to the degree of similarity between targets and distractors and the second being the degree of similarity between distractors themselves. With respect to our study, targets can be regarded as either closed curves or zones. If a target is a zone, as many were in the study, the rectilinear shape of a target zone, as illustrated in Fig. 4, is very similar in shape to the majority of distractor zones, as well as being similar in shape to square closed curves themselves. Therefore, it is not unreasonable to liken the task of identifying a zone in an Euler diagram drawn using squares or rectangles to that of searching for one square/rectangle among many similar looking shapes. These observations regarding shape disciminability are summarised in the 'Shape Discrimination' column of table 5, where the shapes are listed in order of mean performance time.

As squares performed better than ellipses, this suggests that a further graphical property is more influential than shape discrimination. In addition, circles performed better than all other shapes, leading us to seek a graphical property possessed only by diagrams drawn with circles. The principle of contour completion states that shapes made up of smooth curves and exhibit a constant rate of change are easier to interpolate into shapes than those that do not. As we are dealing with 2-dimensional geometric shapes, only circles can support contour completion as a constant rate of change must be exhibited. Thus, this explains why circles outperform all other shapes and ellipses in particular, captured by the 'Contour Completion' column in table 5.

We can explain why squares perform better than ellipses and rectangles by appealing to the principle of proximity. A square has four equal sides at four

	Shape Discrimination	Contour Completion	Highly symmetric
Circle	Y	Y	High
Square	N	N	Medium
Ellipse	Y	N	Low
Rectangle	N	N	Low

Table 5. Perceptual properties of shapes.

equal angles and the proximity of each side, relative to one and other, is constant. The principle of proximity states that the strength of grouping between elements and their properties increases as these elemental properties are brought nearer to each other. Unlike squares (and circles), the shape of an ellipse contravenes the principle of proximity. Following the smooth curvature of an ellipse, starting at an antipodal point of a minor axis, the rate of change of its curvature tends the eye away from its centre until we arrive at the antipodal point of a major axis. Similarly, the shorter sides a rectangle tend away from each other as the eye follows the longer sides. These observations, based on the principle of proximity, are embodied by the fact that ellipses and rectangles exhibit only two lines of symmetry while a square exhibits four and a circle exhibits an infinite number. Whilst there are degrees of symmetry, the last column of table 5 crudely identifies circles and squares as highly symmetric as compared to rectangles and ellipses. The higher degree of symmetry possessed by a circle over a square further supports our statistical results.

We have observed our perceptual sensitivity to the graphical properties of shapes in Euler diagrams. In summary, the smoothness of circles permits effective shape discrimination, they support contour completion, and they are highly symmetric, captured in table 5, distinguishing them from the other three shapes. Further, squares only exhibit high symmetry, distinguishing them from ellipses and rectangles. Table 5, supported by perceptual theory, leads us to posit a further three guides, in an order of priority:

Guide 7 (Shape) *Draw Euler diagrams with circles.*

Guide 8 (Symmetry) *Draw Euler diagrams with highly symmetrical curves.*

Guide 9 (Shape Discrimination) *Draw Euler diagrams so that the zones are discernable from the curves via their shape, but not at the expense of symmetry.*

The first of these guides is strongly supported by the statistical results. The remaining guides require further validation, particularly that for shape discrimination which is only weakly supported by our data. In this case, shape discrimination is the only identified perceptual difference between ellipses and rectangles; we could not conclude that a significant difference exists between these shapes.

7 Threats to Validity

Threats to validity are categorized as internal, construct and external [18]. Internal validity considers whether confounding factors, such as carry-over effect, affects the results and, if so, to what extent. Construct validity examines whether the independent and dependent variables yield an accurate measure to test our hypotheses. External validity considers the extent to which we can generalise the results. The following discusses the primary threats to validity that were considered and addressed to ensure the study is robust and fit for purpose. With regard to internal validity, following two factors were among a number that were considered in an attempt to manage potential disadvantages of our study design:

Carry-over effect: in a repeated measure experiment this threat occurs when the measure of a treatment is effected by the previous treatment. To manage this effect a between group design was employed. Each participant group i.e. square, rectangle, circle and ellipse, was exposed to four different treatments.

Learning effect: the learning effect was considered a threat if participants were not given appropriate training prior to the data collection phase. To reduce this effect, training was given to the participants and they attempted questions used before answering the 18 questions used in the statistical analysis.

Next we consider construct validity by focusing on our dependent variables (error rate, false negatives, and time) and independent variables (diagram and shape), respectively, and examine their rigour for measuring comprehension:

Error rate: all diagrams were drawn to adhere to the six orignal layout guides as well as the layout characteristics detailed above. This drawing approach minimised the possibility of confounding variables creeping into each diagram.

False negatives: to minimise false negatives i.e. a participant selecting the wrong answer while reading it to be the correct answer, the similarity of module and student names was minimised during all phases of the experiment.

Time: to ensure the rigour of time measurements, consideration was paid to the precise duration elapsed interpreting a diagram as well as the units employed to measure time. Further, participants used the same PC located in the same laboratory with no applications running in the background.

Diagram: it was considered a threat if participants did not spend time reading and understanding the diagrams. To manage this threat *diversity* was introduced in the diagrams so that participants had to read and understand each diagram before being able to answer the posed question. It was also considered a threat if the diagrams were regarded as trivial; having only a few curves, zones, or data items was deemed insufficient to yield noticeable differences in response times, should they exist. To manage this, diagrams were designed to exhibit an appropriate level of complexity in order to demand cognitive effort.

Shape: it was essential that the process of drawing equivalent diagrams was carefully planned and executed in order to minimise the threat of unwanted variances between pairs of diagrams.

The following factors consider the limitations of the results and the extent to which the they can be generalised, thus examining their external validity:

Curve shape: Shapes were limited to squares, rectangles, circles and ellipses.

Set theoretic concepts: Euler diagrams conveyed set disjointness, subset and intersecting relationships.

Question styles: three styles of questions were asked: ‘Who’, ‘Which’ and ‘How’.

Participant: participants were representative of a wider student population.

Thus, the results should be taken to be valid within these constraints.

8 Conclusion

The six guides identified in section 2 render effective Euler diagrams. However, irrespective of this guidance, we have observed that the shape of a closed curve significantly affects user comprehension. Consequently, to further improve the effectiveness of an Euler diagram we posited three new guides, focusing on shape, symmetry and shape discrimination. These guides support the common use of circles in both manually drawn Euler diagrams, where they are commonly used [28], and in automated drawing methods such as [24]. Similarly, it may indicate that drawing methods using other shapes, such as rectangles, are less effective. Further, not all data sets can be visualized with Euler diagrams that are both well-formed and drawn using circles. Thus, the two new guides on symmetry and shape discrimination can be employed for drawing diagrams when circles cannot be used.

There is still scope, however, to improve existing drawing methods by gaining further insight into which layout choices lead to more effective diagrams. In particular, there remain numerous graphical properties to which were are known to be perceptually sensitive that have not been given serious consideration when drawing Euler diagrams. As we have demonstrated with shape, the choice of graphical property can have a significant impact on the effectiveness of an Euler diagram. Colour, and its effect on a user's comprehension, is of particular interest when drawing and laying out Euler diagrams. Colour is widely used when automating Euler diagram layout as well as for visualising data. Bertin [4] realised that colour can be used to promote the interpretation of both quantitative and qualitative information. Colour value or lightness is typically prescribed for presenting quantitative information while colour hue is commonly used to represent qualitative information. Consequently, the immediate future direction of our work will be to investigate the best use of colour when visualising data using Euler diagrams.

References

1. B. Alper, N. Riche, G. Ramos, M. Czerwinski. Design study of linesets, a novel set visualisation technique. IEEE Trans. on Visualization and Computer Graphics, 17(12):2259 - 2267, 2011.
2. Architectural Association, London. http://www.aadip9.net/shenfei/, accessed April 2013.
3. F. Benoy P. Rodgers. Evaluating the comprehension of Euler diagrams. In *11th Int. Conf. on Information Visualization*, pages 771–778. IEEE, 2007.
4. J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps.* Uni. of Wisconsin Press, 1983.
5. A. Blake, G. Stapleton, P.J. Rodgers, L. Cheek, J. Howse. Does the orientation of an Euler diagram affect user comprehension? In *18th Int. Conf. on Distributed Multimedia Systems*, pages 185–190. Knowledge Systems Institute, 2012.
6. R. DeChiara, U. Erra, V. Scarano. A system for virtual directories using Euler diagrams. In *Proc. of Euler Diagrams*, vol. 134 of *ENTCS*, pages 33–53, 2005.

7. J. Duncan G. Humphreys. Visual search stimulus similarity. *Psychological Review*, 96:433–458, 1989.
8. M. Eckstein. Visual search: A retrospective. *Journal of Vision*, 11:1–36, 2011.
9. H. Kestler et al. Vennmaster: Area-proportional Euler diagrams for functional GO analysis of microarrays. *BMC Bioinformatics*, 9(1)(67), 2008.
10. G. Farrell, W. Sousa. Repeat victimization and hot spots: The overlap and its implication for crime control and problem-oriented policing. *Crime Prevention Studies*, 12:221–240, 2001.
11. C. Gurr. Effective diagrammatic communication: Syntactic, semantic and pragmatic issues. *J. of Visual Languages and Computing*, 10(4):317–342, 1999.
12. C. Healey, J. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Trans. on Visualisation and Computer Graphics*, 18:1170–1188, 2011.
13. E. Ip. Visualizing multiple regression. *J. of Statistics Education*, 9(1), 2001.
14. P. Isenberg, A. Bezerianos, P. Dragicevic, J. Fekete. A study on dual-scale data charts. In *IEEE Trans. on Visualization and Computer Graphics*, 17(12):2469-2478, 2011.
15. K. Koffka. *Principles of Gestalt Pschology.* Lund Humphries, 1935.
16. L. Micallef, P. Rodgers. Drawing Area-Proportional Venn-3 Diagrams Using Ellipses. Technical Report TR-3-11, School of Computing, University of Kent, 2011. A Java applet is available at http://www.eulerdiagrams.org/eulerAPE.
17. H. Purchase. Which aesthetic has the greatest effect on human understanding? In *5th International Symposium on Graph Drawing*, pages 248–261. Springer, 1997.
18. H. Purchase. *Experimental Human Computer Interaction: A Practical Guide with Visual Examples.* CUP, 2012.
19. N. Riche, T. Dwyer. Untangling Euler diagrams. *IEEE Trans. on Visualisation and Computer Graphics*, 16:1090–1097, 2010.
20. P. Rodgers, L. Zhang, H. Purchase. Wellformedness properties in Euler diagrams: Which should be used? *IEEE Trans. on Visualization and Computer Graphics*, 18(7):1089–1100, 2012.
21. P. Rodgers, G. Stapleton, J. Howse, L. Zhang. Euler graph transformations for Euler diagram layout. In *Visual Languages and Human-Centric Computing*, pages 111–118. IEEE, 2010.
22. P. Simonetto, D. Auber. Visualise undrawable Euler diagrams. In *12th Int. Conf. on Information Visualization*, pages 594–599. IEEE, 2008.
23. P. Simonetto, D. Auber, D. Archambault. Fully automatic visualisation of overlapping sets. *Computer Graphics Forum*, 28(3), 2009.
24. G. Stapleton, J. Flower, P. Rodgers, J. Howse. Automatically drawing Euler diagrams with circles. *J. of Visual Languages and Computing*, 12:163–193, 2012.
25. G. Stapleton, P. Rodgers, J. Howse, L. Zhang. Inductively generating Euler diagrams. *IEEE Trans. of Visualization and Computer Graphics*, 17(1):88–100, 2009.
26. J. Thièvre, M. Viaud, A. Verroust-Blondet. Using Euler diagrams in traditional library environments. In *Euler Diagrams*, vol. 134 of *ENTCS*, pp 189–202, 2005.
27. J. Wagemans et al. A century of gestalt psychology in visual perception: I. perceptual grouping and figure-ground organisation. *Computer Vision, Graphics, and Image Processing*, 31:156–177, 1985.
28. L. Wilkinson. Exact and approximate area-proportional circular Venn and Euler diagrams. *IEEE Trans. on Visualisation and Computer Graphics*, 18:321–330, 2012.