How Should We Use Colour in Euler Diagrams?

Andrew Blake
University of Brighton
a.l.blake@brighton.ac.uk

Gem Stapleton
University of Brighton
g.e.stapleton@brighton.ac.uk

Peter Rodgers
University of Kent
p.j.rodgers@kent.ac.uk

John Howse
University of Brighton
john.howse@brighton.ac.uk

ABSTRACT
This paper addresses the problem of how best to use colour in Euler diagrams. The choice of using coloured curves, rather than black curves, possibly with coloured fill is often made in tools that automatically draw Euler diagrams for information visualization as well as when they are drawn manually. We address the problem by empirically evaluating various different colour treatments: coloured or black curves combined with either no fill or coloured fill. By collecting performance data, we conclude that Euler diagrams with coloured curves and no fill significantly outperform all other colour treatments. Most automated layout algorithms adopt colour fill and are, thus, reducing the effectiveness of the Euler diagrams produced. As Euler diagrams can be used in a multitude of areas, ranging from crime control to social network analysis, our results stand to increase the ability of users to accurately and quickly extract information from their visualizations.

ACM Classification Keywords
H.5.2. User Interfaces: Theory and methods

Author Keywords
Visualization; Euler diagram; colour.

INTRODUCTION
There has been a rapid rise in the volume of data where the data items lie in overlapping sets. Data of this type arise in many situations such as in criminal investigations: sets represent organizations to which people (the data items) belong or locations they frequent. Similar complex data occur in biological settings where data items are genes, whilst sets represent shared features of the genes. Reflecting the importance of understanding such data there have been a large number of techniques proposed for visualizing it such as [1, 10, 19, 23, 26, 27, 30, 31, 34]. Most of these techniques represent the sets using Euler diagrams.

Euler diagrams are often regarded a natural [28] and intuitive [25] way to depict sets. They represent sets using graphical elements called closed curves. The interior of each curve represents items that are in the set [24, 28]. Figure 1 shows a simple classification. The curve labelled 'Mammals' intersects with the curve labelled 'Aquatic' meaning that there are some mammals that are aquatic. 'Cats' is contained within 'Mammals' meaning that all cats are mammals. 'Cats' is disjoint from 'Aquatic' so there are no cats that are aquatic.

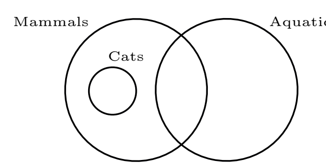

Figure 1. Simple Euler diagram.

Euler diagrams are used in a large variety of application areas including the natural sciences [12], art and architecture [2], education [16], criminology [13], computer file organisation [11] and classification systems [32]. Given such abundant use, there is a strong motivation to understand how choices made when drawing Euler diagrams will impact upon users' comprehension.

This paper focuses on the use of colour in Euler diagrams, informed by how colour has been used in a variety of automated layout techniques. Flower and Howse [14], who produced the first work on automated Euler diagram drawing, use black curves with no fill. Stapleton et al.'s method employs coloured curves with no fill [30] whereas Kestler et al.'s VennMaster [18] uses black curves with colour fill. By far the predominant choice amongst developers of these layout techniques, however, is coloured curves with colour fill [9, 10, 20, 23, 33, 29].

By providing insight into the relative strengths of different colour treatments, we will be better positioned

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
VINCI '14, August 05 - 08 2014, Sydney, NSW, Australia
Copyright 2014 ACM 978-1-4503-2765-7/14/08…$15.00.
http://dx.doi.org/10.1145/2636240.2636838

to develop improved visualization methods. Our first contribution is an empirical comparison of four colour treatments: black curves and no fill, coloured curves and no fill, black curves and colour fill, and coloured curves and colour fill; when using colour, each set being visualized is assigned a unique colour. We tested the hypothesis that the use of colour would aid comprehension. We discovered that using Euler diagrams with coloured curves and no fill significantly outperformed the other three colour treatments in terms of accuracy. This significant difference was more pronounced when the diagrams exhibited topological properties known to hinder comprehension. For the other three colour treatments, no significant differences where exhibited in the accuracy data. Moreover, there were no significant differences when considering completion time. We conclude that using coloured curves with no fill will best aid performance. 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/colour/colour.html`. As with all empirical studies, including those cited in the next section, our results are valid within the constraints of the study.

RELATED WORK: EULER DIAGRAM LAYOUT CHOICES

There are numerous choices that need to be made when drawing Euler diagrams for a given data set. We categorize these choices into three types: descriptional (the abstract syntax level), topological and graphical (both at the concrete syntax level). The first choice that must be made is descriptional, which determines the *zones* that must be present in the Euler diagram [30]. The zones are the smallest regions in Euler diagrams which represent overlaps between the sets; Figure 1 has four zones inside its three curves. A descriptional choice determines which zones are present, in addition to those that represent non-empty overlaps between sets. An example is given in Figure 2, where the lefthand diagram is well-matched whereas the righthand diagram contains additional (shaded) zones.

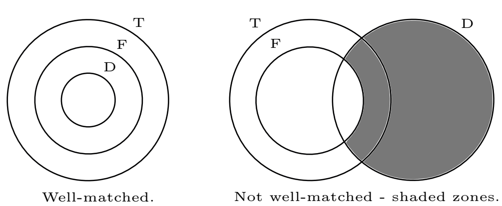

Figure 2. Choices of abstract syntax.

This choice is embodied in Gurr's theory of well-matchedness [15]:

> "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."

Euler diagrams are well-matched precisely when they do not include any 'extra' zones. 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 of an Euler diagram. At the concrete syntax (i.e. drawn diagram) level, a variety of topological choice are known to impact comprehension [24]. These include curves running concurrently and points at which three curves meet; a complete discussion list, with examples, is presented later as they form an important aspect of our study. Diagrams that possess such properties are said to be not well-formed and the results of [24] are summarised here:

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 when ascertaining an effective layout of an Euler diagram. Benoy and Rodgers [3], with their work on aesthetics, also acknowledge the importance of making correct graphical choices when laying out 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.*

There are many other graphical choices that might be considered. Bertin [4] identifies both planar and retinal variables, which constitute a variety of graphical choices, to which we are known to be perceptually sensitive. With respect to planar variables, Blake et al. 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, both of which are fundamental choices that must be made when drawing Euler diagrams. Recently Blake et al. established that shape significantly impacts users' comprehension [6], leading to three further guides:

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.*

With respect to shape, we have observed the significant impact a retinal variable can impose on users' comprehension of Euler diagrams. Consequently, our motivation to address the question "how does the use of colour in Euler diagrams affect users' comprehension?" is further reinforced. Presently, the research community does not know whether or how colour choice impacts the comprehension of Euler diagrams.

EXPERIMENTAL DESIGN

As we are interested in the impact of colour choice on the comprehension of Euler diagrams, our study requires a range of diagrams to be drawn adopting a variety of different colour treatments. For the purposes of this study, and congruent with previous studies [1, 5, 17, 20, 21, 22, 23], we view comprehension in terms of task performance: one diagram is more comprehensible than another diagram if users can interpret it, on average, more accurately or more quickly.

Colour Treatments

We compare four colour treatments listed below using a between group design. Figures 3 to 6 illustrate the colour treatments listed below and represent four Euler diagrams used during the study (from this point forward, all Euler diagrams in the paper are scaled versions of those used in the study):

1. B&N: black curve and no fill (figure 3),
2. C&N: colour curve and no fill (figure 4),
3. B&F: black curve and colour fill (figure 5), and
4. C&F: colour curve and colour fill (figure 6).

Bertin, when prescribing the application of retinal variables, recommends colour for data classification [4]. Thus, we hypothesize that coloured curves will outperform black curves, more so when diagrams are not well-formed. However, the impact of colour fill versus no fill is less clear.

As we are studying the effect of colour, the choice of colours used in our diagrams is a core design feature of our study. The colours used were chosen from the palette in figure 7. To help ensure the colours in the palette were uniformly distinct from each other we adopted the approach of Brewer et al [7]. Brewer prescribes maintaining, as far as is possible, even intervals between colour hue while keeping both saturation and lightness constant. We will refer to a value range between 0 and 255 and specify the colours in terms of their hue, saturation, lightness (*HSL*) and alpha channel. Hue values were evenly incremented by 32, starting with a value of 32 to create the palette of 8 colours in figure 7. Saturation and lightness were set to values of 197 and 171, respectively, with the exception of the green hues, 2 and 3 as illustrated in figure 7. The green hue represents a large proportion of the colour spectrum. Consequently, the saturation for colours 2 and 3 were set to 124 and 197 repetitively. The alpha channel was set to 255 for the curve colours and 100 for the coloured fill. Below, we explain how these colour and fill option are applied to diagrams used in our study.

Drawing Diagrams

To reflect our hypotheses, a mixture of well-formed and non-well-formed diagrams were needed for the study. We drew 36 diagrams, of which 18 were well-formed (WF; e.g. figures 3 to 6) and, therefore, 18 were non-well-formed (NWF; e.g. figures 8 to 13). Euler diagrams are NWF if they exhibit any one of the following topological properties: a brushing point, which occurs when two curves meet but do not cross (figure 8); concurrency, which occurs when two or more curve segments share, or partially share, the same line (figure 9); duplicate curves labels, which occur when two or more curves share the same label (an example of which is 'Peru' in figure 10); a disconnected zone, which occurs when a zone consists of one or more minimal regions(an example of which appears in both 'Rwanda' and 'Chad' in figure 11); a triple point, which occurs when three or more curves pass through a particular point (figure 12); a non-simple curve, a curve that self-intersects (figure 1). Diagrams that have one or more of these properties may benefit from the use of colour as it can be hard to identify the curves.

The diagrams were drawn sensitive to the layout guidelines presented earlier, except that NWF diagrams deliberately contravened guide 2; necessarily, this contravention implies some other guides need not be met such as that curves need not be circular. Also, the diagrams in the study adhered to the following conventions, to ensure they 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. all closed curves had a 3 pixel stroke width,
4. all diagrams were drawn in an area of 810×765 pixels,
5. the curve labels were written using upper case letters in Times New Roman, 14 point size, font in bold,
6. data items were written using lowercase letters, except that the first letter was capitalised, and with Arial 12 point size font,
7. each curve label was positioned closest to its corresponding curve, and

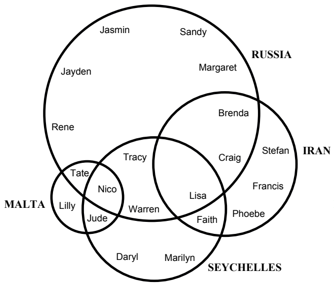

Figure 3. Type 1 B&N.

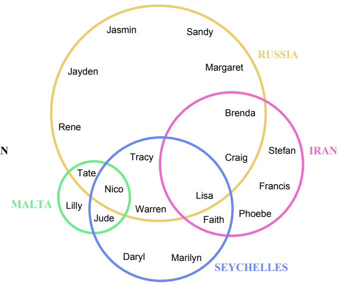

Figure 4. Type 1 C&N.

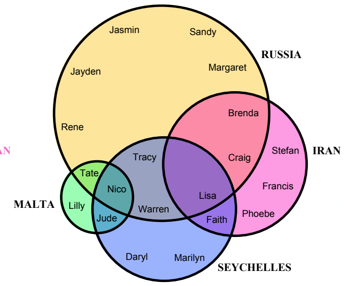

Figure 5. Type 1 B&F.

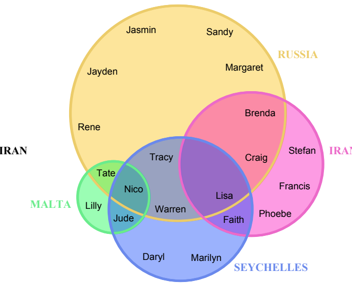

Figure 6. Type 1 C&F.

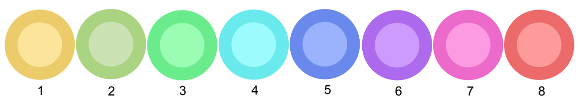

Figure 7. Colour palette.

8. data items were evenly distributed within each zone.

These conventions were informed by previous studies [6, 5] and validated by a pilot study. NWF diagrams that exhibited either concurrent curves as illustrated in figure 9, of which there were three, or non-simple curves as illustrated in figure 13, of which there were also three, did not precisely adhere to convention 2. Their topologies necessitated that some curves' scale deviated from the above convention.

Diagram Types

The diagrams drawn were required to exhibit variety to enable reasonably general conclusions to be reached, in so far as is possible with empirical research. To this end, all diagrams were drawn pertaining to the following three characteristic types:

1. Type 1: 4 sets (figure 3 to 6). When curve labels are unique, the four sets were represented by 1 large curve, 2 medium curves and 1 small curve. In these diagrams were 11 zones and 20 data items. In the one instance where the labels were not unique, there are five curves, 1 large, 2 medium and 2 small. In this case there were 12 zones and 20 data items.

2. Type 2: 6 sets (figure 14). When curve labels are unique, the six sets were represented by 1 large curve, 4 medium curves and 1 small curve. In these diagrams were 15 zones and 30 data items. In the one instance where the labels were not unique, there eight curves, 1 large, 4 medium and 3 small. In this case there were also 15 zones and 30 data items.

3. Type 3: 8 sets (figure 15). When curve labels are unique, the eight sets were represented by 2 large curves, 4 medium curves and 2 small curves. In these diagrams were 19 zones and 40 data items In the one instance where the labels were not unique, there 11 curves, 2 large, 2 medium and 5 small. In this case there were also 19 zones and 40 data items.

The largest number of sets was chosen to be eight because the resulting diagrams exhibited a level of complexity observed in real world examples and is consistent with previous studies [6, 24]. Moreover, to guarantee the colours were perceptually distinct the palette was chosen to be limited to eight colours.

Twelve diagrams were drawn of each type, split equally between WF and NWF. Colours were assigned to types as follows, referencing the colour palette in figure 7:

1. Type 1 diagrams were treated with colours 1, 3, 5 and 7.

2. Type 2 diagrams were treated with colours 1, 2, 4, 5, 7, 8.

3. Type 3 diagrams were treated with colours 1 to 8.

In addition, colours were randomly applied to curves and curve labels had the same colour as their corresponding curves. In NWF diagrams, multiple curves representing the same set (i.e. when duplicate curve labels were present), had the same colour treatment. When curves had a colour fill, as the alpha channel was to 100, the zones are filled with the combined colours of the curves they are inside.

Question Styles

Congruent with [1, 20, 24], all diagrams conveyed 'real-world' information as it was regarded pertinent to the reader. Further, abstract representations were considered a barrier to the participants understanding. Therefore, we chose to visualize information about people and the countries they have visited. This information was regarded generic so all participants should be equally familiar with this context. Country names were derived from a variety of continents. People names were first names only and a mixture of both male and female names, and reflected a variety of ethnicities. Following [24], and similar to [1, 20], three styles of question were specified, 'Who', 'Which' and 'How many', that allowed us to elicit the following type of information:

1. Who has visited RUSSIA, IRAN and SEYCHELLES?

2. Which country has been visited by 6 people?

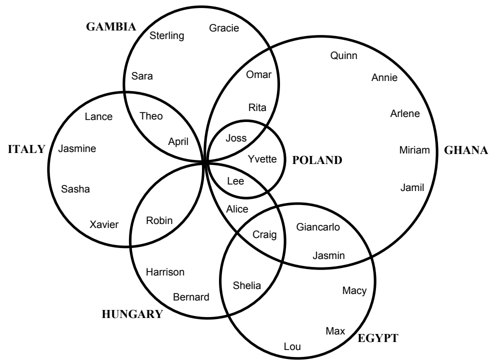

Figure 8. Type 2 NWF with brushing points.

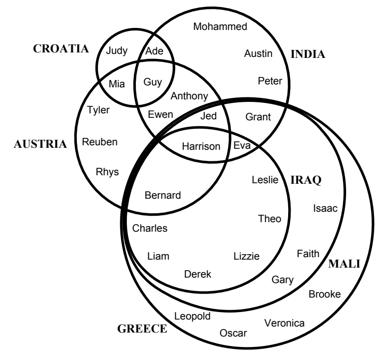

Figure 9. Type 2 NWF with partial concurrency.

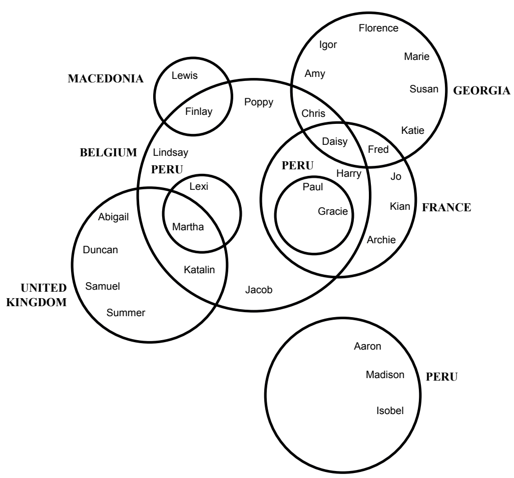

Figure 10. Type 2 NWF with duplicate curve labels.

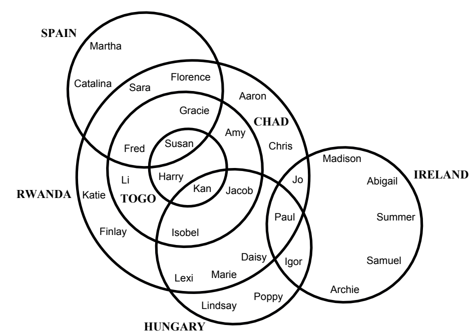

Figure 11. Type 2 NWF with disconnected zones.

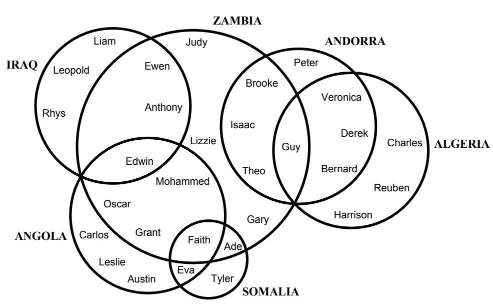

Figure 12. Type 2 NWF with triple points.

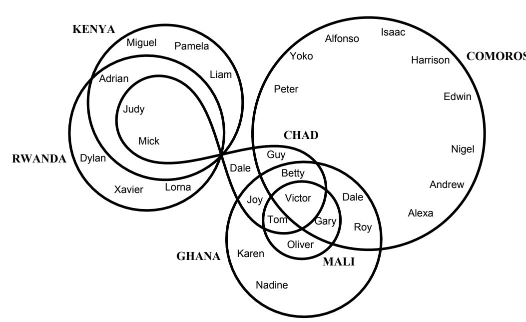

Figure 13. Type 2 NWF with a non-simple curve.

3. How many people have visited both MALAWI and SAN MARINO?

The above questions were asked of figures 3, 14 and 15 respectively. During the study, questions had a choice of four or five possible answers depending on the type of question and the number of sets represented in the diagram. In each instance, only one answer was correct.

EXPERIMENT EXECUTION

Initially, a pilot study was conducted consisting of eight participants, two per group. This was executed with out any issues and so the data could be carried forward for analysis. A further 72 participants were recruited for the main study. All 80 participants (68 M (1 of which was colour blind), 12 F, ages 18 to 32) were randomly allocated to equal sized groups. They were all students from the University of Brighton's School of Computing, Engineering and Mathematics and they spanned both undergraduate and postgraduate levels.

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 36 diagrams were presented in a random order.

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. To ensure colours were optimally displayed the monitor provided a 1920 x 1200 resolution, 0.270 mm pixel pitch, 300 cd/m^2 brightness and a contrast ratio of 1000:1 (static). 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 a canteen voucher worth £6 for their contribution to the research.

The first phase of the experiment was initial training. All participants were asked whether they were familiar with Euler diagrams. Some responded that they had seen Venn diagrams but none acknowledged familiarity with Euler diagrams. Consequently, all participants were treated as having no previous experience of Euler diagrams and were 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.

STATISTICAL ANALYSIS

The 80 participants in the main study each answered 36 questions giving a total of 2880 observation. We consider error data to be a more important indicator of performance in our evaluation. This is because the time taken to complete a task does not matter if ultimately the answer is incorrect, consistent with Alper et al. [1]. Results are regarded as significant if $p \leq 0.05$.

Error Analysis

Of the 2880 observations there were a total of 128 errors (error rate: 4.4%). The error counts and accuracy rates for colour treatment for all diagrams, as well as well-formed and non-well-formed diagrams, are presented in table 1. The accuracy rates are also presented in figure 16.

To establish if there was an overall effect we performed a χ^2 goodness-of-fit test. Overall, there were significant differences between the colour treatments, with a p-value of 0.002. Hence, we conclude that colour treatment impacts on user comprehension when interpreting Euler diagrams. In order to rank the treatments we then proceeded to perform pairwise comparisons. Comparing colour curve and no fill (C&N) with each of the other treatments in turn yielded significant differences in all cases (largest p-value 0.016). However, there were no other significant differences between the remaining pairs of treatments. As colour curve and no fill yielded the fewest number of errors our hypothesis that coloured curves aid comprehension is supported.

Analysing the error data for the well-formed case did not reveal any significant difference between the colour treatments. However, in the non-well-formed case there are significant differences ($p = 0.005$). Again, comparing colour curve and no fill (C&N) with each of the other treatments in turn revealed significant differences in all cases (largest p-value 0.040). As with the well-formed case, there were no other significant differences between the remaining pairs of treatments. As we hypothesised, the use of colour curves is more beneficial

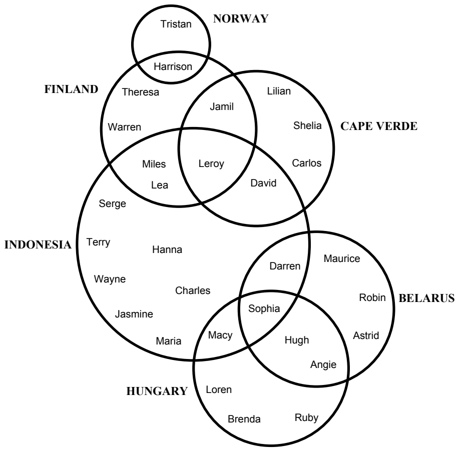

Figure 14. Type 2 B&N.

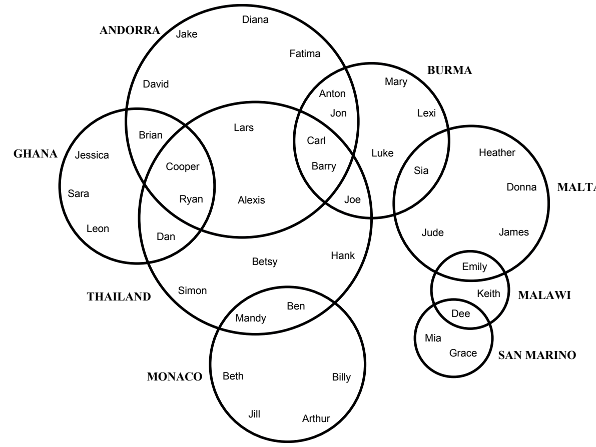

Figure 15. Type 3 B&N.

	Overall	B&N	C&N	B&F	C&F	p-value
All	$\frac{128}{2880} = 4.4\%$	$\frac{42}{720} = 5.8\%$	$\frac{15}{720} = 2.1\%$	$\frac{40}{720} = 5.6\%$	$\frac{31}{720} = 4.3\%$	0.002
Well-formed	$\frac{53}{1440} = 3.7\%$	$\frac{15}{360} = 4.2\%$	$\frac{8}{360} = 2.2\%$	$\frac{16}{360} = 4.4\%$	$\frac{14}{360} = 3.9\%$	0.395
Non well-formed	$\frac{75}{1440} = 5.2\%$	$\frac{27}{360} = 7.5\%$	$\frac{7}{360} = 1.9\%$	$\frac{24}{360} = 6.6\%$	$\frac{17}{360} = 4.7\%$	0.005

Table 1. Summary of error data.

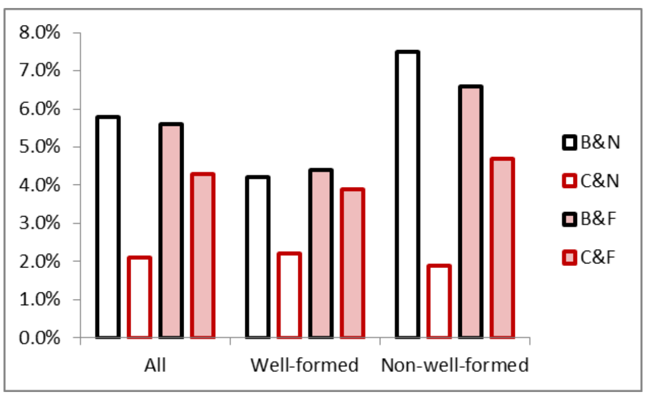

Figure 16. Accuracy data.

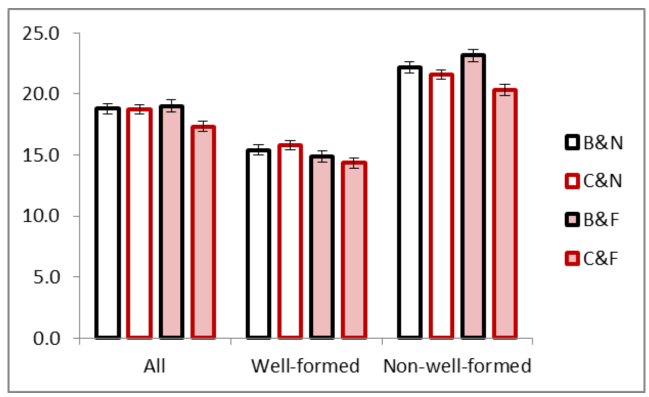

Figure 17. Time data.

in the non-well-formed case.

Time Analysis

Of the 2880 observations the grand mean was 18.45 seconds (sd: 11.77) and means across colour treatments are given in table 2 (standard deviations in brackets), and illustrated in figure 17.

To determine if there existed significant variation across colour treatments, we performed a Repeated Measures Analysis of Variance (RM-ANOVA). Participants were regarded to be a random factor in the analytical model. This is a conservative approach to the analysis as the RM-ANOVA did not estimate the effect of each of the participants in the sample, but instead estimated the variability attributable to the participants. The RM-ANOVA was performed on logged data to achieve normality. There was no significant effect of colour treatment on the time taken to answer questions ($F(3, 35) = 0.60$, $p = 0.614$). Breaking down into well-formed and non-well-formed also yields no significant differences ($F(3, 17) = 0.66$, $p = 0.578$; $F(3, 17) = 0.69$, $p = 0.560$ respectively). Hence, we conclude that colour treatment does not significantly impact on time taken.

Summary of Performance Data

To summarise the analysis of the performance data, we present evidence that suggests colour treatment does impact upon user comprehension of Euler dia-

	Overall	B&N	C&N	B&F	C&F	p-value
All task	18.45 (11.77)	18.78 (11.50)	18.69 (10.14)	19.00 (13.34)	17.33 (11.84)	$p = 0.614$
Well-formed	15.09 (8.65)	15.40 (8.93)	15.77 (7.71)	14.86 (8.92)	14.33 (8.94)	$p = 0.578$
Non-well-formed	21.81 (13.41)	22.16 (12.74)	21.61 (11.37)	23.15 (15.57)	20.33 (13.52)	$p = 0.560$

Table 2. Summary of time data.

grams. While colour treatment does not significantly impact upon time taken to comprehend Euler diagrams it does significantly impact upon the number of errors accrued. Euler diagrams with black curves and no fill, (B&N), yield the most errors while Euler diagrams with coloured curves and no fill, (C&N), yield the fewest errors. The effect is largest when the diagrams are not well-formed.

INTERPRETATION OF RESULTS

To interpret the results we first remind the reader that an Euler diagram represents sets using closed curves. The interior of each closed curve contains the data that are in a set. Thus, each closed curve defines the boundary between one set and another. The primary task of users in the study is to identify data within closed curves. Users must visually segregate the boundaries defined by one closed curve from another, in order to interpret relationships between closed curves and identify data therein.

Hue is favoured over shape when segregating boundaries [8]. Diagrams with black curves and no fill are solely reliant on users identifying the shape of the curve while diagrams with colour curves and no fill have the additional advantage of hue. This helps explain why diagrams with coloured curves and no fill performed significantly better than diagrams with black curves and no fill, which yielded most errors. However, boundary segregation is impaired when a "secondary, irrelevant dimension varies" [8]. Diagrams with colour fill contained random variations of hue as a consequence of the relationships between closed curves. We regard colour fill as the 'secondary, irrelevant dimension', thus further explaining why diagrams with colour curves and no fill performed significantly better than diagrams with colour fill.

Consequently, the results of our empirical study, supported by the underlying visual perception theory, leads us to posit a further guide:

Guide 10 (Colour). *Draw Euler diagrams with curves that have no fill and different colours for each set represented.*

To summarise, coloured curves with no fill significantly out perform all other colour treatments. Our results are, in part, supported by the work of [8]. The implication of our results is far reaching. Not only do they provide a robust foundation for guidance regarding colour, but the colour treatments identified in the study are a direct reflection of those employed by current researchers and practitioners. Of particular note is that the optimal colour treatment is not that adopted by many layout techniques, such as [9], [10], [18], [23], [29], [33].

THREATS TO VALIDITY

Threats to validity are categorized as internal, construct and external [22]. 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, the 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 affected by the previous treatment. To manage this effect a between group design was employed. Each participant group, i.e. B&N, C&N, B&F and C&F, was exposed to one of the treatments.

Learning effect: the learning effect was considered a threat if questions were delivered in a fixed order. Therefore, questions were delivered to each participant in a random order. Further, to minimise the learning curve during the main study, participants were given appropriate training prior to the data collection phase.

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

Error rate: diagrams were drawn to adhere to our layout guides with the exception of non-well-formed diagrams. All diagrams adhere to generic layout characteristics. 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 country and people 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.

Colour: strict convention was adhered to when treating diagrams with colour.

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

Colour palette: there were eight colours with properties based on the work of [7].

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

Question styles: three styles of questions were asked: 'Who', 'Which' and 'How many'.

Participant: participants were representative of a wider student population.

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

CONCLUSION

In this paper we evaluated the use of colour in Euler diagrams, focusing on both curve colour and curve fill. We performed a controlled experiment to compare four different colour treatments by considering task performance. In terms of accuracy, using coloured curves with no fill significantly outperformed all other treatments. No significant differences were observed in timing data. Key results are summarised as follows:

1. Overall, coloured curves with no colour fill outperformed black curves either with or without colour fill and coloured curves with coloured fill, leading to a new layout guide: *use a different coloured curve for each set represented, and do not use coloured fill.*

2. For non-well-formed Euler diagrams, using different coloured curves with no fill brings more benefit than in the general case.

3. The majority of automated Euler diagram layout techniques adopt both coloured curves and coloured fill, contravening the new layout guide of avoiding colour fill.

Perceptual theory has aided our explanation of these results: colour fill acts as a distracter from the primary task of identifying the curves.

We are very interested in exploring possible interactions between all of the Euler diagram layout guides with a view of prescribing a hierarchy for their application especially as it is, in general, not possible to ensure that they all hold. Further, we would be interested to observe whether the effect of symmetry (guide 8) is further enhanced or diminished with the application of colour (guide 10). Consequently, the immediate future direction of our work will be to investigate the interaction between the guidelines for drawing Euler diagrams. As a result, we will be better placed to use Euler diagrams for visualizing information.

It will also be interesting to explore the impact of colour in a richer set of diagrams. For instance, the number of sets is currently limited to the eight and we did not consider area proportionality in our study. Moreover, our study only used three curve sizes. As Euler diagrams become more complex, we hypothesise that the use of colour will be even more significant provided the colours are still perceptually distinct. Other choices we made, such as line thickness and font size, could also be varied and empirically tested but there is no evidence to suggest that such changes will impact comprehension. Lastly, it will be interesting to explore layout choices in a dynamic or interactive setting which has yet to be considered.

REFERENCES

1. B. Alper, N. Riche, G. Ramos, and M. Czerwinski. Design study of linesets, a novel set visualisation technique. *IEEE Transactions on Visualization and Computer Graphics*, 17:2259–2267, 2011.

2. Architectural Association, London. http://www.aadip9.net/shenfei/, acc. April 2013.

3. F. Benoy and 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, and J. Howse. Does the orientation of an Euler diagram affect user comprehension? In *18th Int. Conf. on Distributed Multimedia Systems, Visual Languages and Computing*, pages 185–190. Knowledge Systems Institute, 2012.

6. A. Blake, G. Stapleton, P.J. Rodgers, L. Cheek, and J. Howse. Does the shape of an Euler

diagram's closed curve affect user comprehension? In *Diagrams*, pages 124-138, Springer, 2014.

7. C.A Brewer. http://www.colourbrewer2.org, accessed October 2013.
8. C. Callaghan. Interference and dominance in texture segregation: Hue, geometric form, and line orientation. *Perception & Psychophysics*, 46(4):299–311, 1989.
9. S. Chow and F. Ruskey. Drawing area-proportional Venn and Euler diagrams. In *Graph Drawing 2003*, pages 466–477. Springer, 2003.
10. C. Collins, G. Penn, and M. Sheelagh T. Carpendale. Bubble sets: Revealing set relations with isocontours over existing visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1009–1016, 2009.
11. R. DeChiara, U. Erra, and V. Scarano. A system for virtual directories using Euler diagrams. In *Euler Diagrams*, ENTCS, pages 33–53, 2005.
12. H. Kestler et al. Vennmaster: Area-proportional Euler diagrams for functional GO analysis of microarrays. *BMC Bioinformatics*, 9(1)(67), January 2008.
13. G. Farrell and 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.
14. J. Flower and J. Howse. Generating Euler diagrams. In *Diagrams*, pages 61–75, Springer, 2002.
15. C. Gurr. Effective diagrammatic communication: Syntactic, semantic and pragmatic issues. *Journal of Visual Languages and Computing*, 10(4):317–342, 1999.
16. E. Ip. Visualizing multiple regression. *Journal of Statistics Education*, 9(1), 2001.
17. P. Isenberg, A. Bezerianos, P. Dragicevic, and J. Fekete. A study on dual-scale data charts. In *IEEE Trans. on Visualization and Computer Graphics*, pages 2469 – 2478. IEEE, 2011.
18. H. Kestler, A. Muller, T. Gress, and M. Buchholz. Generalized Venn diagrams: A new method for visualizing complex genetic set relations. *Journal of Bioinformatics*, 21(8):1592–1595, 2005.
19. H. Kestler, A. Muller, H. Liu, D. Kane, B. Zeeberg, and J. Weinstein. Euler diagrams for visualizing annotated gene expression data. In *Euler Diagrams 2005*, 2005.
20. W. Meulemans, N. Henry Riche, B. Speckmann, B. Alper, and T. Dwyer. Kelpfusion: A hybrid set visualization technique. *IEEE Transactions on Visualization and Computer Graphics*, 19(11):1846–1858, 2013.
21. H. Puchase. Which aesthetic has the greatest effect on human understanding? In *Graph Drawing*, pages 248–261. Springer, 1997.
22. H.C Purchase. *Experimental Human Computer Interaction: A Practical Guide with Visual Examples*. Cambridge University Press, 2012.
23. N. Riche and T. Dwyer. Untangling Euler diagrams. *IEEE Transactions on Visualisation and Computer Graphics*, 16:1090–1097, 2010.
24. P. Rodgers, , L. Zhang, and H. Purchase. Wellformedness properties in Euler diagrams: Which should be used? *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1089–1100, 2012.
25. P. Rodgers, G. Stapleton, J. Howse, and L. Zhang. Euler graph transformations for Euler diagram layout. In *Visual Languages and Human-Centric Computing* , pages 111–118. IEEE, 2010.
26. P. Rodgers, L. Zhang, and A. Fish. General Euler diagram generation. In *Diagrams*, pages 27–31. Springer, 2008.
27. P. Simonetto. *Visualisation of Overlapping Sets and Clusters with Euler Diagrams*. PhD thesis, Université Bordeaux, 2012.
28. P. Simonetto and D. Auber. Visualise undrawable Euler diagrams. In *12th Int. Conf. on Information Visualization*, pages 594–599. IEEE, 2008.
29. P. Simonetto, D. Auber, and D. Archambault. Fully automatic visualisation of overlapping sets. *Computer Graphics Forum*, 28(3), 2009.
30. G. Stapleton, J. Flower, P. Rodgers, and J. Howse. Automatically drawing Euler diagrams with circles. *Journal of Visual Languages and Computing*, 12:163–193, 2012.
31. G. Stapleton, P. Rodgers, J. Howse, and L. Zhang. Inductively generating Euler diagrams. *IEEE Transactions of Visualization and Computer Graphics*, 17(1):88–100, 2009.
32. J. Thièvre, M. Viaud, and A. Verroust-Blondet. Using Euler diagrams in traditional library environments. In *Euler Diagrams*, ENTCS, pages 189–202, 2005.
33. J. Vihrovs, K. Prūsis, K. Freivalds, P. Ručevskis, and V. Krebs. An inverse distance-based potential field function for overlapping point set visualization. *Int. Conf. on Information Visualization Theory and Applications*, pages 29-38, 2014.
34. L. Wilkinson. Exact and approximate area-proportional circular Venn and Euler diagrams. *IEEE Transactions on Visualisation and Computer Graphics*, 18:321–330, 2012.