

Wellformedness Properties in Euler Diagrams: Which Should be Used?

Peter Rodgers, Leishi Zhang, and Helen Purchase

Abstract— Euler diagrams are often used to visualize intersecting data sets in applications such as criminology; genetics, medicine and computer file systems. One interesting aspect of these diagrams is that some data sets cannot be drawn without breaking one or more “wellformedness properties”, which are considered to reduce the user comprehension of the diagrams. However, it is possible to draw the same data with different diagrams, each of which breaks different wellformedness properties. Hence, some properties are “swappable”, so motivating the study of which of the alternatives would be best to use. This paper reports on the two empirical studies to determine how wellformedness properties affect comprehension. One study was with abstract data, the other was with concrete data that visualized students’ enrollment on university modules. We have results from both studies that imply that diagrams with concurrency or disconnected zones perform less well than other some other properties. Further, we have no results that imply that diagrams with brushing points adversely affect performance. Our data also indicates that non-simple curves are preferred less than diagrams with other properties. These results will inform both human diagram designers and the developers of automated drawing systems on the best way to visualize data using Euler diagrams.

Index Terms— Euler diagrams, Venn diagrams, empirical studies, information visualization.

1 INTRODUCTION

xxxx-xxxx/0x/\$xx.00 © 200x IEEE

Euler diagrams [6] are widely used in information visualization because of the intuitive way in which they represent set intersections. In many application areas, items are placed inside the intersections of the curves in the diagrams, which allows users to understand which sets the item belongs to, and which it does not. There are a number of uses for Euler diagrams, including: criminal data analysis [7]; genetics [12]; medicine [17]; and extending file system capabilities [5]. Fig. 1 shows two Euler diagrams, inspired by that from [11]. Both visually describe the same relationship between parts of the British Isles; each has the same underlying data, but they are drawn differently. Here, territories are items inside curves labelled with UK legal terminology. For example, it can be seen that “Northern Ireland” is part of the “UK” and is on the island of “Ireland”, whilst “Wales” is also part of the “UK” but is not on “Ireland”.

Euler diagrams and Venn diagrams [20] are often confused. Venn diagrams must contain every possible set intersection, but Euler diagrams may omit some intersections. Hence, every Venn diagram is an Euler diagram, but not every Euler diagram is a Venn diagram.

- Peter Rodgers is with the University of Kent, UK. E-mail: P.J.Rodgers@kent.ac.uk.
- Leishi Zhang is with the University of Konstanz, Germany E-mail: leishi.zhang@uni-konstanz.de.
- Helen Purchase is with the University of Glasgow, UK. E-mail: helen.purchase@glasgow.ac.uk.

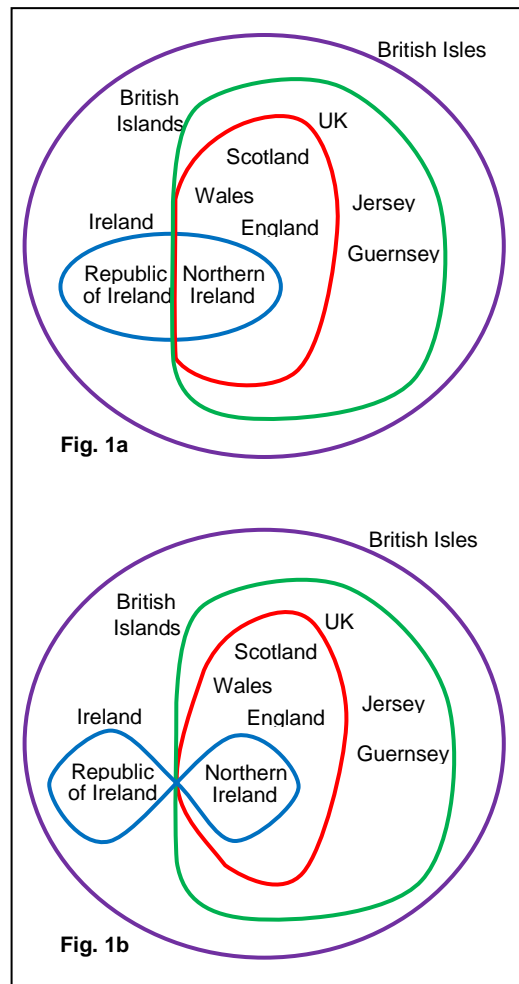


Fig. 1. Two alternative “British Isles” Euler diagrams.

We define an Euler diagram to be a collection of labelled closed curves. Curve labels may occur more than once in a diagram. A connected component of the plane that is bordered by curve segments is termed a minimal region. More than one minimal region may be inside the same set of curves, and, for a particular set of curves, the set of all such minimal regions is called a zone, which can be described by the curve labels it is inside. For example, the diagram given in Fig. 2 has curves with labels **A**, **B** and **C** and zones \emptyset , $\{A\}$, $\{B\}$, $\{C\}$, $\{A,C\}$, $\{B,C\}$, $\{A,B,C\}$. Note that the zone $\{C\}$ is composed of two minimal regions. Note also that zone $\{A,B\}$ is not present, which would be required if this was a Venn diagram with three curves.

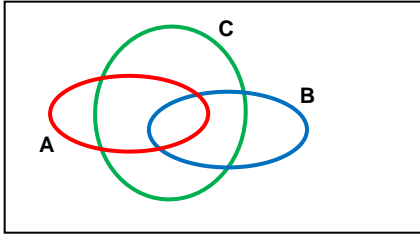


Fig. 2 . An example Euler diagram.

An important aspect of drawing Euler diagrams is that of wellformedness. The notion was introduced by Flower and Howse [10] and has been subsequently formalized [18]. Wellformedness properties relate to relationships between curves and regions in the diagram.

In this paper we are concerned with six of the most common properties. Diagrams that break these properties are shown in Fig. 3a-f, all of which have the same zones present. The six wellformedness properties considered in this work are:

- **Brushing Point (BP):** there are two or more curves that meet but do not cross at a point. In Fig. 3a there are two pairs of curves that meet at a brushing point, one pair is “ALGORITHMS” and “GAMES PROGRAMMING”.
- **Concurrency (CC):** two or more curve segments are concurrent. Fig. 3b has two concurrent sections, one concurrent section involves the curves labelled “DYNAMIC WEB” and “VISUAL PROGRAMMING”. To show concurrency, we have separated the curve segments slightly so that they run next to each other, rather than sharing exactly the same route.
- **Duplicated curve label (DC):** two or more curves have the same label. In Fig. 3c, there are three curves with label “DATABASES”.
- **Disconnected zone (DZ):** one or more zones consists of more than one minimal region. In Fig. 3d, there are two disconnected zones, one is the zone that is inside the curve “INFORMATION SYSTEMS” and no other curves, which consists of two minimal regions.
- **Non-simple curve (NS):** a curve self-intersects. In Fig. 3e there are two non-simple curves, one is the curve “CONCURRENCY”.
- **Triple point (TP):** three or more curves cross at the same point. In Fig. 3f there are two triple points, one

is at the intersection of the curves “NETWORKING”, “PROJECT” and “SOFTWARE SECURITY”.

Fig. 3g shows a wellformed diagram, breaking no wellformedness properties.

These wellformedness properties are thought to adversely affect the understanding of a diagram for two main reasons: either they increase the local density of information at a particular point (concurrency, non-simple curves, triple points, brushing points), or they split a single entity into multiple, spatially distinct sections (disconnected zones and duplicated curve labels). In the first case there is a greater cognitive effort in distinguishing between curves in the diagram. In the second case, it is easy to miss one part of a disconnected entity, and in addition, maintaining the location of all parts of the entity in short-term memory requires increased cognitive effort.

Many data sets cannot be drawn without breaking some wellformedness property. Fig. 1 shows such a data set. However, there is usually a choice between the properties that are broken, as seen in Fig. 1 where there is a choice between concurrency (Fig. 1a) and non-simple curves (Fig. 1b). In addition, general automatic Euler diagram generation tools often break some wellformedness properties even when diagrams can be drawn in a wellformed manner, including those described in [4][15][16][21]. Hence, there is a strong motivation for discovering which properties most affect users understanding when examining Euler diagrams.

The research question we ask here is “Which wellformedness properties are the most preferable to break?”. To answer this we have conducted two empirical studies to examine the relative usability of wellformedness properties.

In the initial study we presented users with alternative versions of Euler diagrams which must be drawn with at least one wellformedness property broken. We added items to different regions of the diagrams, and asked the participants questions to gauge their understanding of the abstract data represented by the diagrams.

The second study was prompted by two aspects of the first study. Firstly, we felt that examining a concrete context might yield additional interesting results. Secondly, the diagrams in the initial study often had more than one wellformedness property broken, which made individual analysis of properties difficult. This was a result of using diagrams that could only be drawn in a non-wellformed manner. A further motivation for the second study was recent independent work [9] reported after we completed our initial study. As with our initial work, it examined abstract data with no real world instantiation, but has contradictory findings. This may be because of the differences in between their tasks and diagrams and ours. They looked at tasks related to logic based interpretation of diagrams, with no individual items in the zones, but used shading to indicate empty zones.

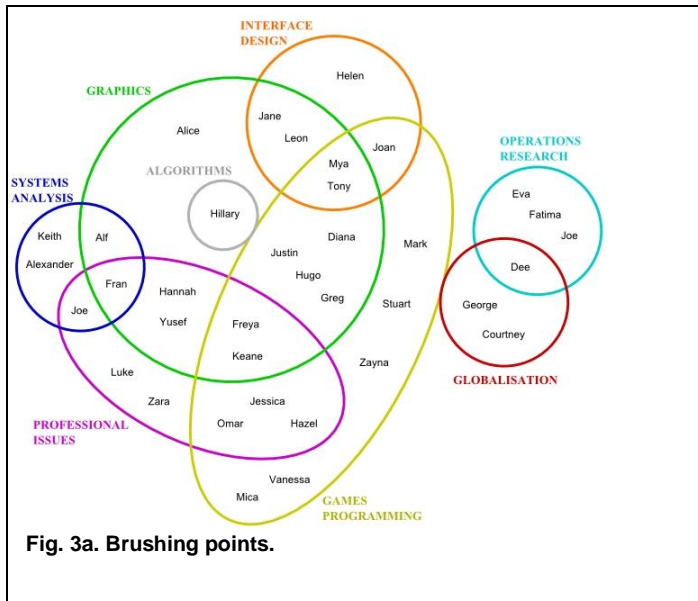


Fig. 3a. Brushing points.

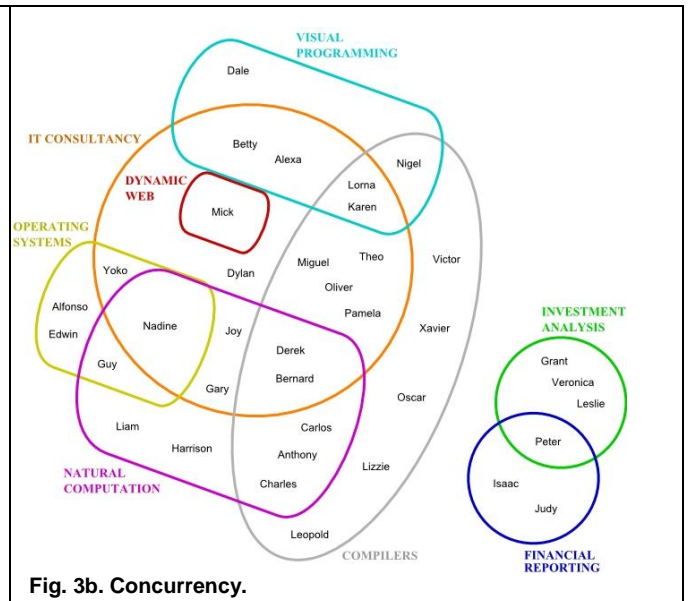


Fig. 3b. Concurrency.

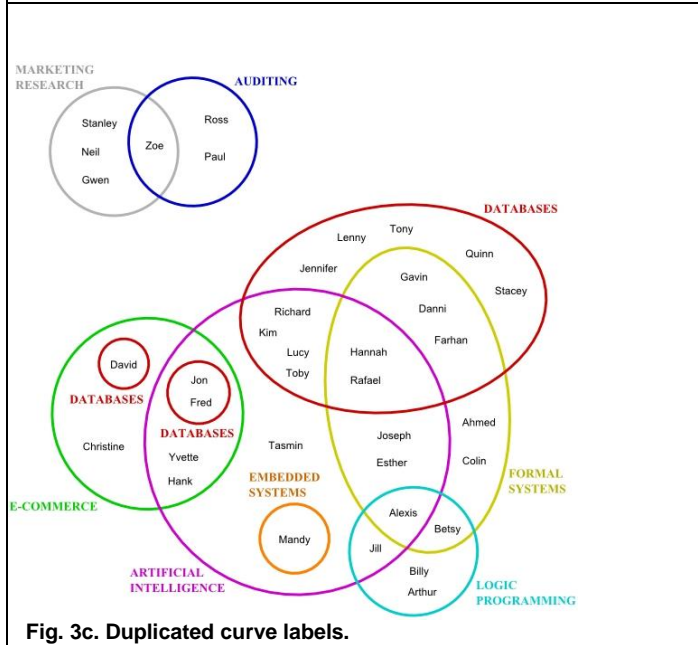


Fig. 3c. Duplicated curve labels.

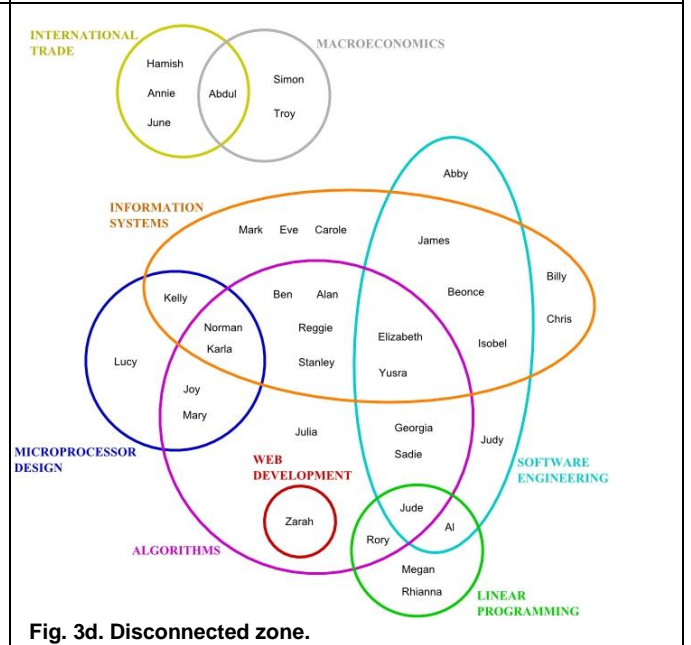


Fig. 3d. Disconnected zone.

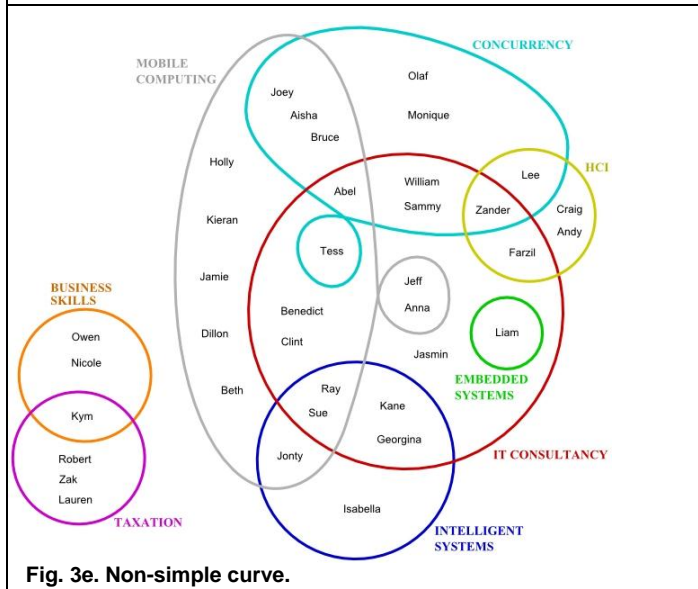


Fig. 3e. Non-simple curve.

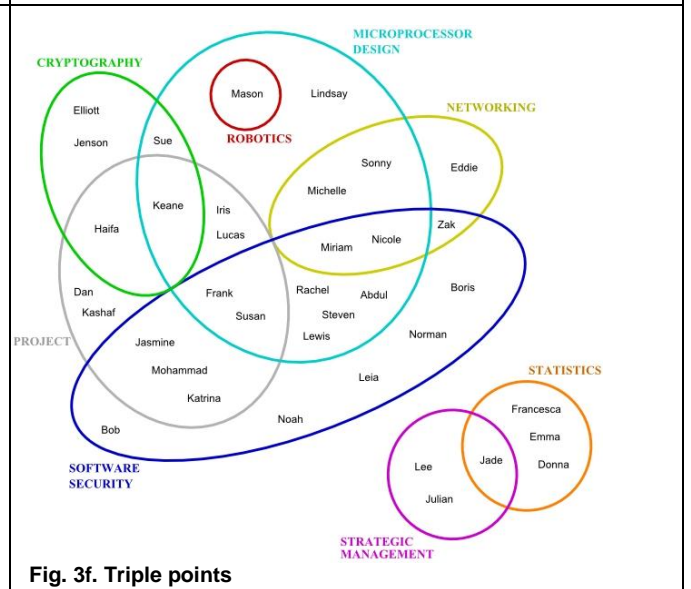
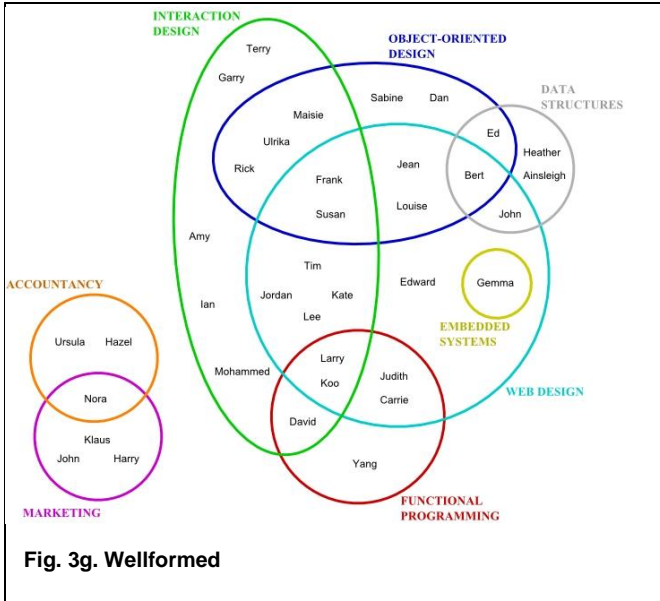


Fig. 3f. Triple points



These issues strongly motivated a follow on, application centred study in order to further explore these issues. For our application area, we chose to visualize the modules students were attending at a university. As our participants would be students at the University of Kent, we could consider them experts in this type of data. The diagrams created for the study were based on wellformed diagrams that were manipulated so that one property was broken at a time. See figures 3a-g for example diagrams from the study.

The results from both the initial study and the second study imply that tasks on diagrams with disconnected zones and concurrency take longer than on diagrams with other properties broken. Disconnected zones were worse than concurrency. The second study strongly indicates that duplicated curve labels negatively affect understanding. Further, the results of both suggest that diagrams with brushing points do not adversely affect understanding, and they are not significantly worse than wellformed diagrams.

The importance of the results in this paper is that we can now give scientifically based guidance to people who draw Euler diagrams. In addition, the developers of automated drawing methods now have guidelines for which properties they should prioritize. For example, our results suggest that tasks take longer to perform on diagrams with concurrency than on diagrams with non-simple curves, so implying that Fig. 1b is a better design than Fig. 1a. However, from our preference data, a diagram with concurrency is preferred by users over a diagram with non-simple curves, so possibly reversing the quality of the diagrams if user preference is the primary motivation of the designer.

These findings contradict the work of other researchers [9], as their conclusions were that brushing points and triple points were the properties that most affected understanding adversely, however, our studies show that brushing points have no negative impact on understanding and triple points are not the most important property. They

also conclude that concurrency can actually aid understanding, in fact we found that concurrency has a significantly adverse effect. There are many possible explanations for these differing conclusions, including the differences in diagram construction and the nature of the tasks. But our work implies that, in some application areas at least, concurrency should be avoided and that brushing points are not a problem for user understanding.

In terms of other empirical work examining Euler diagrams, the layout of curves, rather than wellformedness has been examined [2]. An initial study indicated that wellformedness properties have an impact on user understanding [8]. In addition, a study has examined alternative representations for Euler diagrams and graphs embedded in them [13].

More generally, the use of Euler diagrams as a visualization method for grouping items is supported by the preattentive processing concept of closure [19]. Euler and similar diagrams are also considered to aid inference, using the concept of ‘free-ride’ [1], where adding a curve can allow the deduction of information not present in either the original diagram or added curve. The use of visual structure has been discussed [22], which may have implications for the design of Euler diagrams. For instance, suggesting that round shapes have a more organic feel than squares which seem to imply isolation.

The remainder of this paper is structured as follows: in Section 2 we outline the initial study; Section 3 details the second study; and finally, Section 4 presents our conclusions and discusses possible future work.

2 INITIAL STUDY: ABSTRACT DATA

Here we consider the first study, which was aimed at abstract data that has to be drawn in a non-wellformed manner.

This study involved the identification of zones in a diagram. As discussed in Section 1, a zone is the set of minimal regions that are contained by the same curve labels. The set of zones for a diagram is called the abstract description of the diagram. For example, the diagrams given in Fig. 4 all have the abstract description $\{\emptyset, \{A\}, \{B\}, \{C\}, \{A,B\}, \{A,C\}, \{B,C\}, \{B,D\}, \{C,D\}, \{A,B,C\}\}$, where \emptyset is the infinite outer zone (the region not contained by any curves) that must always be present. In order to simplify the notation, we often write such abstract descriptions as follows: **A B C AB AC BC BD CD ABC**.

2.1 Initial Study Experiment Design

To allow for sensible questions to be posed in the study, we used Euler diagrams with items inside, which corresponds to the use of Euler diagrams in a number of application areas, and is very different from the mathematically oriented diagrams presented in another study [9]. The diagrams were kept simple, so labelling was restricted to curves, as identifying zones relies on knowing the label of the curve that the zone is contained within. The diagrams had three or four curves. See Figures 4 and 5 for example diagrams used in the study.

We took the six wellformedness properties described in Section 1 and chose several abstract descriptions that must be drawn with one or more property broken. We represented them in two or more ways, each of which showed the breaking of different wellformedness properties. An example of different ways of drawing the same abstract description is shown in Fig. 4, where Fig. 4a exhibits concurrency, Fig. 4b exhibits a duplicated curve label and Fig. 4c exhibits a non-simple curve combined with a triple point. An example of drawing a diagram in two ways is shown in Fig. 5 where a diagram is drawn with a triple point in Fig. 5a and a duplicated zone in Fig. 5b (the outer zone is present both in the border of the diagram and in the middle of the diagram).

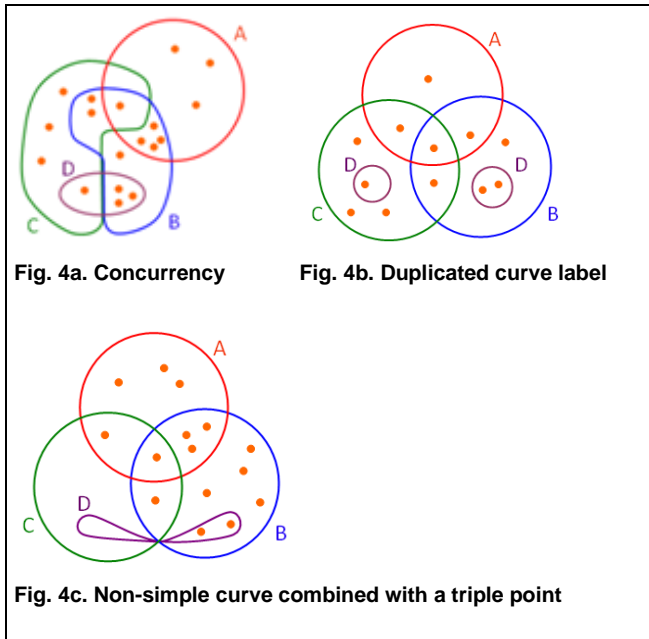


Fig. 4. Three ways of drawing the abstract description A B C AB AC BC BD CD ABC.

In many cases only one wellformedness property was broken, however, in some diagrams two needed to be broken. This is because, for data sets that must be drawn in a non-wellformed manner, two of the wellformedness properties (brushing points and non-simple curves) always require others to also be broken; a diagram that is drawn with a brushing point either includes a triple point or a non-simple curve. A diagram that is drawn with a non-simple curve either includes a triple point or brushing point. Hence, we could not test the usability of diagrams on this data without sometimes considering an additional property.

The diagrams were hand drawn. When creating the diagrams we concentrated on ensuring that each was as understandable as we could reasonably make it. This meant, where possible, using familiar shapes for curves which are known to be popular when visualizing Euler diagrams, such as circles and ellipses. Where such shapes were not feasible, smooth and unambiguous shapes were used. In the diagrams, the curves were designed to intersect in a way that avoided ambiguity in curve routing,

so at the intersection point each curve followed an uninterrupted continuous path. The exception to smooth curves and continuous paths at intersections is in some cases where curve segments were concurrent. In these cases a straight line and an abrupt change in direction when the curve segments separated was considered to be most usable, as it highlights the transition between concurrent and non-concurrent sections. We used different colours for each curve in the diagrams, because colour is typically applied when displaying Euler diagrams in real world use.

We chose two types of question: type 1 asks “how many items are there in a particular zone” and type 2 “asks which zone in the diagram contains a specific number of items”. The question of type 1 for the diagram in Fig. 6 was “How many items(s) are there in zone C?”, with the answer “3”. The question of type 2 was “Which zone contains 2 items?” with the answer “ACD”. These questions had the benefits of being applicable to a simple type of Euler diagram, could only be answered by understanding the zones of the diagram (so therefore the participants would demonstrate some understanding of the diagram), and had multiple choice answers. We developed 72 main questions that were based on 36 diagrams, with each diagram having two questions associated with it: one of type 1, and the other of type 2. The 36 diagrams were in 14 groups, consisting of two or three diagrams, and each diagram in a group had the same abstract description but exhibited the breaking of a different wellformedness property. The groups are shown in Table 1, with the codes for the wellformedness properties given next to the property descriptions in Section 1.

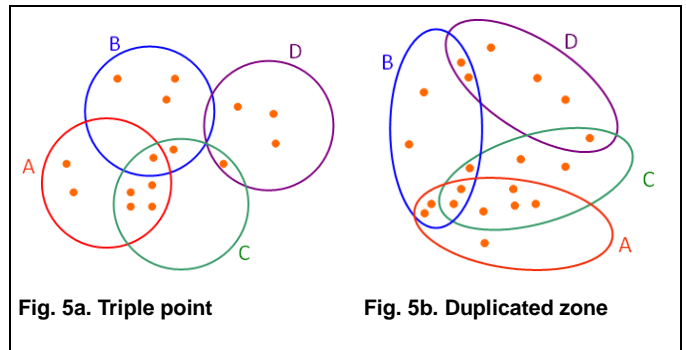


Fig. 5. Two ways of drawing the abstract description A B C D AB AC BC BD CD ABC.

During the formation of the study we conducted two pilot studies to tune the experimental process, questions and timing. As a result the time given to answer the questions was reduced from one minute to 20 seconds, and the explanation of the tasks was made clearer. The first pilot consisted of four participants, the second of five participants. The data from the pilots was discarded.

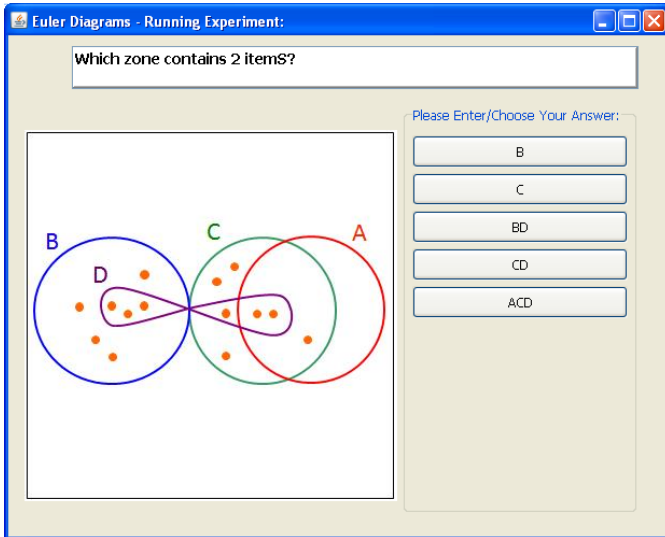


Fig. 6. Initial study software screenshot.

For the main study we tested 21 participants, all undergraduates studying computing related degrees at the University of Kent. The tests were conducted by the same investigator, who was alone and uninterrupted with the participant. All participants were asked the same 72 main questions, but which were presented to them in a random order. Nine training questions were presented to them before the main questions, which were always in the same order, but for which the results were not considered in the statistical analysis. The training questions and the random order for the main questions were designed to mitigate the learning effect. When participants were tested, they were first read an introductory script that included an explanation of Euler diagrams and terms such as ‘zone’. The first part of the test then followed. This was computer based. Here participants were presented with multiple choice questions that had a time limit. The window in which the diagram and question appeared filled the majority of a 19” screen. See Fig. 6 for an example of how the software presented such a question. Once a question was answered the next was presented. In the main study they were limited to 20 seconds per question. We limited the time allowed for answering a question to encourage participants to answer quickly. Limiting the time was an attempt to put the participants under pressure and so they would be more likely to answer incorrectly. Shortly before the time limit was reached a warning beep was sounded. If the time limit was reached before an answer was given the next question would be presented. After every nine questions there was a break. This gave participants a rest from the test for around a minute. We recorded the time taken, whether the answer was correct and whether the question was answered within the time limit.

Once all 72 main questions were completed, the second part of the study commenced. The participants were read another script and given a questionnaire that asked them some details about their background. It also included a free text section so that they could add qualitative information about the tasks in the study. They were then asked to rank six wellformedness pictures for preference.

Table 1. Initial study test diagrams

Abstract Description	Zones	Properties Broken
A B C AC BC ABC	6	TP CC DZ
A B AB AC BC ABC	6	TP DZ
A B AC BC	4	NS/BP CC DC
A B C AB ABC	5	NS/TP CC DC
A B AB AC BC	5	NS/TP CC DC
A B C D AB AC BC BD CD ABC	10	TP DZ
A B C D AB AC BC BD ABC BCD	10	TP DZ
A B C AB AC BC BD CD ABC BCD	10	TP DZ
A B C AC BD CD ACD	7	NS/BP CC DC
A B C D AB BC BCD	7	NS/TP CC DC
A B C AB AC BC BD CD ABC	9	NS/TP CC DC
A B C AB AC AD ABD ACD	8	BP/TP CC
A B C D AB AC BD CD	8	BP/TP CC
A B C D AB AC BD CD ABD	9	BP/TP CC DC

After the participants had completed the questionnaire, they were given a debrief sheet to explain the purpose of the study and given five UK Pounds for their contribution to the research.

2.2 Initial Study Results

Here we discuss the quantitative feedback for the initial study. We asked qualitative and preference questions. The results of the qualitative response showed no interesting patterns, so have not been presented. Regarding the preference question, it is not clear how accurate the results are, as multiple properties appeared on some diagrams shown, hence we have not included these results.

2.2.1 Initial Study Timing Results

Firstly, we looked to see if there any difference in performance over all the different wellformedness properties. Fig. 7 shows the time taken to complete the tasks. Our response time analysis includes data from questions which were answered incorrectly as response time is intended to be an indication of cognitive effort required, independent of whether this cognitive effort resulted in a correct answer or not. For those tasks where an answer was not given with the 20 seconds time limit, the response time was recorded as 20 seconds.

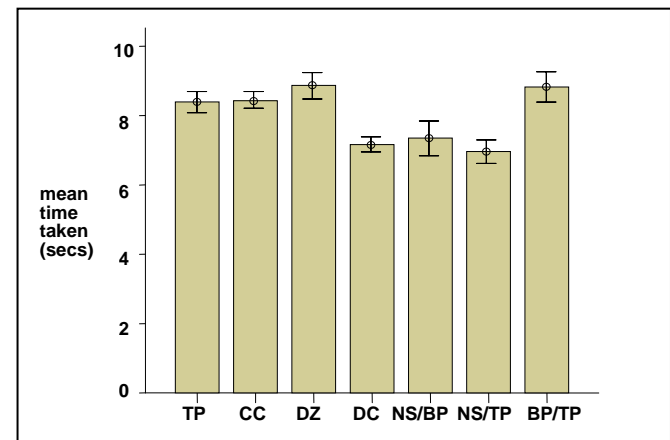


Fig. 7. Initial study average time taken for each wellformedness property. The error bars show the standard error.

A repeated measures ANOVA test revealed that there is a significant effect of property on response time ($F=7.517 > F(df=6, 120; p=0.05) = 2.18$). A post-hoc Tukey test reveals where the differences lie, shown in Table 2 (this table shows only significant results).

Table 2. Initial study time differences

		Mean Difference (Secs)	Significance
CC	DC	1.27	0.002
CC	NS/BP	1.47	0.017
DZ	DC	1.71	0.001
DZ	NS/BP	1.52	0.020
DZ	NS/TP	1.91	0.010

Thus we have evidence that:

- tasks on diagrams with concurrency take longer than tasks on diagrams with duplicated curve labels or non-simple curves combined with brushing points;
- tasks on diagrams with disconnected zones take longer than tasks on diagrams with duplicated curve labels or diagrams with non-simple curves combined with either brushing points or triple points.

2.2.2 Initial Study Accuracy Results

With regard to errors, (i.e. the number of incorrect or incomplete answers over all tests), the number of correct answers out of the 1512 tests was 1348, giving an overall error rate of 0.122. Fig. 8 shows the error rates for the wellformedness properties.

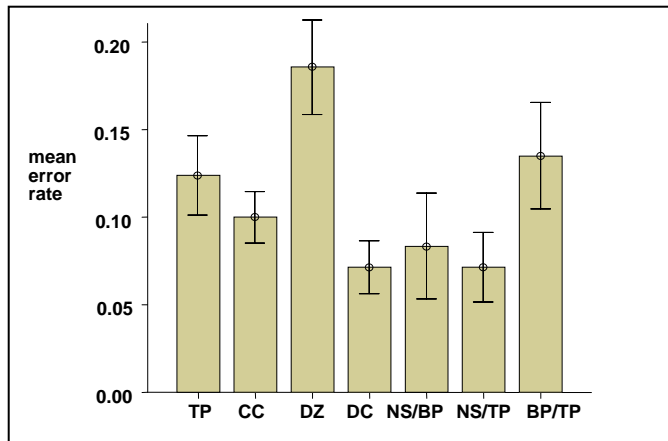


Fig. 8. Initial study average error rate for each wellformedness property.

A repeated measures ANOVA test revealed that there is a significant effect of property on response time ($F=3.460 > F(df=6, 120; p=0.05) = 2.18$). A post-hoc Tukey test reveals where the differences lie, as shown in Table 3 (again only the significant result is shown).

Table 3. Initial study error differences

		Mean Difference	Significance
DZ	DC	0.114	0.005

2.3 Initial Study Analysis

The results relating to time taken in Section 2.2.1 show some interesting patterns. In particular, it takes longer to complete tasks involving diagrams with concurrency compared to some other properties (duplicated curve labels or non-simple curves combined with brushing points), as shown in Table 2, which was not expected by the investigators. However, this might be explained by the non-regular nature of the curves when concurrency is present whereas, when other wellformedness properties are present, circles and ovals can often be used to present all of the curves in the diagram. This is illustrated in Fig. 9 where the concurrent version of the diagram must be drawn with non-regular shapes, whereas the duplicated curve label version can be drawn with circles. This might imply that the familiarity of users with such shapes is an important factor in diagram comprehension. Alternatively, it is necessary for participants to derive the segments of a particular curve when it enters and leaves the concurrent section. This may introduce an extra cognitive load, so explaining the extra time taken.

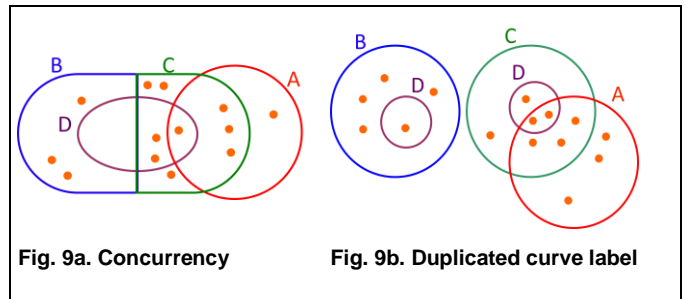


Fig. 9. Two ways of drawing the abstract description A B C AC BD CD ACD.

The other result relating to time was more expected. This is that tasks on diagrams with disconnected zones take longer than diagrams with duplicated curve labels or diagrams with non-simple curves combined with either brushing points or triple points. A possible reason for this is that representing the same zone in a disconnected manner means that the one of the features of Euler diagrams that is considered to be useful has been lost: that single entities are connected so all items can be seen in the same group. Grouping by closure has been identified as a preattentive feature [19].

In terms of error rates, Section 2.2.2, the data is less conclusive than that of the time data, probably due to the low overall error rates in the test. The one result (see Table 3), that tasks on diagrams with disconnected zones are more difficult than diagrams with duplicated curve labels confirms the similar result derived from the time data.

3 SECOND STUDY: APPLICATION BASED

After the first study, it was clear that a follow on study with a concrete application area would allow the discovery of more interesting, and perhaps more relevant, results. The initial study used data that had to be drawn in a non-

wellformed manner. This was problematic because the properties often needed to be combined, so direct comparison between individual wellformedness properties was not possible and so generating useful preference data was not sensible. In the second study, we avoided combining of properties by using data that did not have to be drawn with properties broken. This also meant that getting preference data for particular properties would be feasible.

3.1 Second Study Experiment Design

For the second study we chose the concrete application area of visualizing students taking modules on a university degree course. This is a concept that should be familiar to all participants in the study, as they were drawn from students studying computing at the University of Kent. The diagrams were based on more complex data sets than the initial study, and were generated by first developing a wellformed diagram, then drawing six different versions. All seven versions had the same underlying data, but each had a different wellformedness property broken. Each of the seven versions (six with one property broken plus one wellformed version) was used 3 times, but rotated (pseudo-randomly, with at least 45 degree angle between the three diagrams with the same property) and different labels applied, see figures 3a-g for seven such diagrams. Unlike the initial study, we made sure all the diagrams were of similar complexity to avoid this as an additional confounding factor. The diagrams all had eight curves, one with 19 zones and 37 students, the second with 17 zones and 34 students, and the third with 21 zones and 35 students.

As with the initial study, each diagram was hand drawn with an attempt to make the visualization as comprehensible as possible. Here, due to the larger size of the diagrams, we ensured that the non-wellformed diagrams had two instances of the particular property broken. As before, we used different colours for each curve in the diagrams.

There were three types of question used: “Which”, “How” and “Who”, all had 5 multiple choice answers, with the correct answer always being unique. Here are examples of each type of question:

1. “Which module is being taken by 6 students?”, for Fig. 3a. Answer: “INTERFACE DESIGN”;
2. “How many students are taking both IT CONSULTANCY and MOBILE COMPUTING but not CONCURRENCY”, for Fig. 3e. Answer: “4”;
3. Who is taking both LINEAR PROGRAMMING and SOFTWARE ENGINEERING but not ALGORITHMS?”, for Fig. 3d. Answer: “A1”.

The “Which” question required examining the contents of an entire curve or set of curves with the same label. The “How” question required participants to examine the intersection between several curves, and find the union of multiple zones. The “Who” question required the discovery of a single zone.

After conducting a pilot study with 5 participants, we adjusted some of the diagrams, and increased the number of questions. The data from the pilot was discarded.

There were 21 different diagrams formed from combining the seven wellformedness properties with the three abstract descriptions. For each of these we developed a question with the question targeting part of the diagram where the property occurred (of course, for the wellformed diagram, this targeting was not possible, and the location was chosen arbitrarily). To avoid participants learning to immediately look in the vicinity of the property when seeking a solution, we then took each of these 21 diagrams, and rotated them, changed the curve colours and renamed all the labels, twice. These further 42 diagrams were assigned the same question (modulo relabeling) as one of the original 21 and are non-targeted because the question did not relate to the wellformedness property in the diagram. This gave a total of 63 questions for the main study.

To illustrate the notion of a targeted question, question 2 in the above list is a targeted question as “MOBILE COMPUTING” and “CONCURRENCY” have non-simple curves in Fig. 3e. In contrast, question 1 is a non-targeted question as “INTERFACE DESIGN” is not a curve involved in either of the brushing points of Fig 4a.

For the main study we tested 22 participants, all undergraduates studying computing related degrees at the University of Kent. The tests were conducted by one of two investigators. The investigator was alone and uninterrupted with the participant. All participants were asked the same 63 main questions, but which were presented to them in a random order. Fourteen training questions were presented to them before the main questions, which were always in the same order, but for which the results were not considered in the statistical analysis. The training questions and the random order for the main questions were designed to mitigate the learning effect. When participants were tested, they were first read an introductory script that included an explanation of the task with two examples. The first part of the test then followed. This was computer based.

Here participants were presented with multiple choice questions that were not time limited, although in the introduction they were told that their time taken to perform the test would be recorded. The window in which the diagram and question appeared in most of a 19” screen. See Fig. 10 for an example of how the software presented a question. Once a question was answered the next was presented. No time limit was given, as the pilot study indicated that subjects did not take an excessive time to answer any of the questions. From our experience with the initial experiment, a high success rate was expected, and we intended to analyse performance based on time taken. After every nine questions there was a break. This gave participants a rest from the test for around 20 seconds. For each diagram, we recorded the time period between display of the diagram and submission of the answer, and whether the answer was correct.

Once all 63 main questions were completed, the second part of the study commenced. The participants were read another script and given a questionnaire that asked them some details about their background. It also included a free text section so that they could add qualitative information about the tasks in the study. They were then asked to rank seven example wellformedness pictures for preference (these diagrams are shown in figures 3a-g).

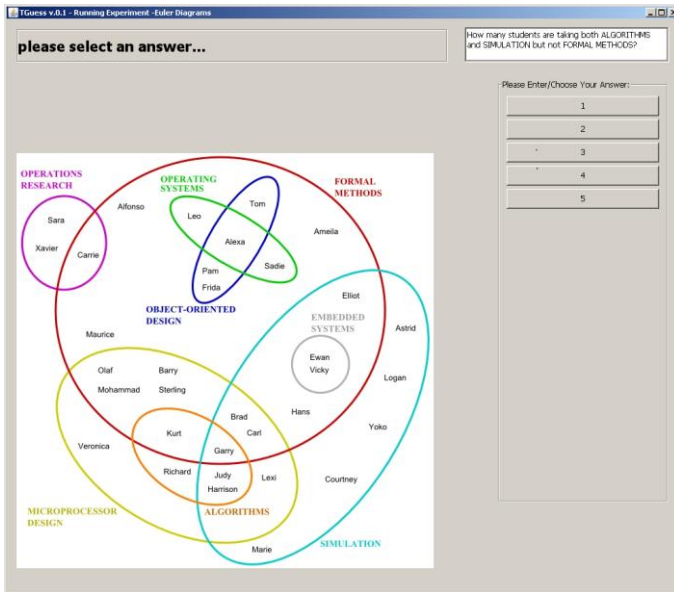


Fig. 10. Second study software screenshot.

Before they saw the diagrams for the preference ranking they were asked to write down answers text in response to these two questions:

- “What made answering the questions easy?”
- “What made answering the questions hard?”

After they had seen the diagrams and ranked them according to preference they were asked to write down text in response to the question:

- “Now you have examined some diagrams, are there any features of them that made some of the questions particularly easier or more difficult than others? Why?”

After the participants had completed the questionnaire, they were given a debrief sheet to explain the purpose of the study and given five UK Pounds for their contribution to the research.

3.2 Second Study Results

Here we discuss the quantitative feedback for the second study. We present the preference data after the performance data.

As expected, the error rate was low: over all 1386 data points 1304 were correct and 82 were incorrect. There was no variability in the error data according to condition, so we concentrate on the timing results for the quantitative data.

The qualitative feedback produced little of interest, except that four participants queried the colours used in

the study, and thought that either the grey or yellow shades were too light.

3.2.1 Timing Results

The timing data was not normally distributed, so a non-parametric analysis method was used.

Firstly, we looked to see if there any difference in performance for diagrams where the question was targeted at the area of the diagram where the wellformedness property was evident. Due to a non-symmetric application of questions to diagrams (one diagram was mistakenly given an additional non-targeted question in place of that question for another diagram), we aggregated using means.

We are interested in the question “which of the well-formed properties, when broken, performs worst?”, and more specifically:

- “when questions targeted at zones affected by the broken property are asked?”;
- “when other questions are asked?”;
- “overall?”.

For a), we looked at the data points when questions targeted at zones affected by the broken property are asked (i.e. one related question), see Fig. 11 for the results.

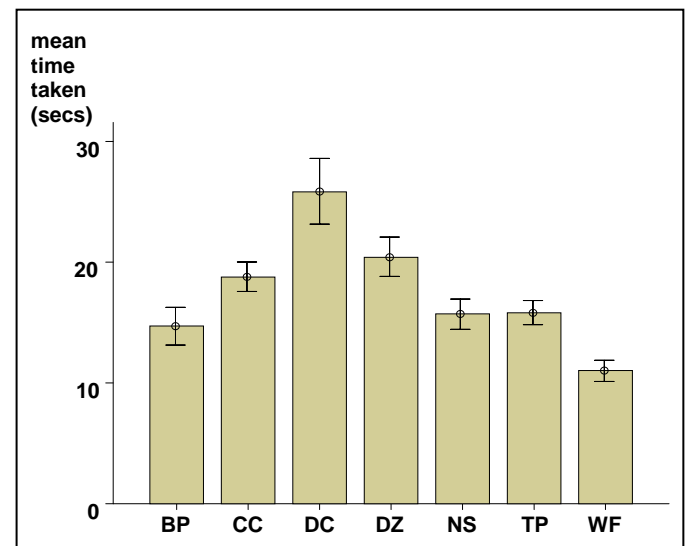


Fig. 11. Second study performance on targeted questions.

A Friedman non-parametric related-measures test produced a Chi-square value of 68.14. As this is greater than chi-square ($p=0.05, 6$) = 12.59, there are significant differences between the properties. A post-hoc Nemenyi pair-wise comparison test revealed where the significant differences lie. Hence, in the case of targeted cases, we have evidence that:

- tasks on diagrams with duplicated curve labels take a longer time than tasks on diagrams with all the other properties;
- tasks on diagrams with all the other properties apart from brushing points take a longer time than tasks on wellformed diagrams;

- tasks on diagrams with either concurrency or disconnected zones take a longer time than tasks on diagrams with brushing points.

For b), the case when non-targeted questions are asked (i.e. two non-related questions), see Fig. 12 for the results.

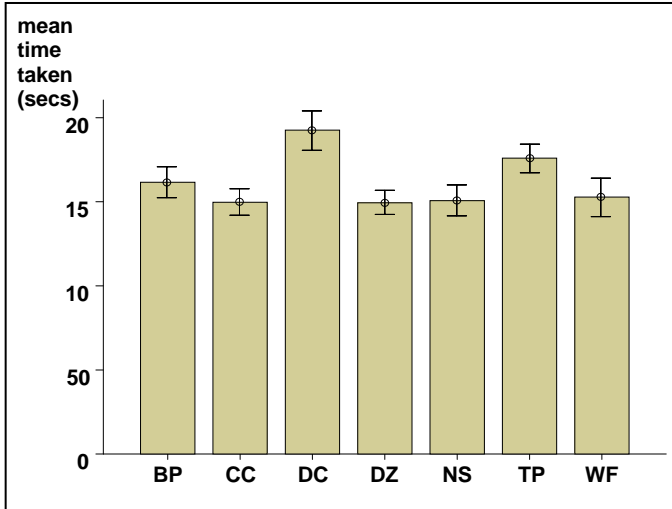


Fig. 12. Second study performance on non-targeted questions.

A Friedman non-parametric related-measures test produced a Chi-square value of 40.01. A post-hoc Nemenyi pair-wise comparison test revealed where the significant differences lie. Hence, in the case of non-targeted cases, we have evidence that:

- tasks on diagrams with duplicated curve labels take longer time than tasks on diagrams with all the other properties apart from triple points;
- tasks on diagrams with triple points take a longer time than tasks on diagrams with either disconnected zones, concurrency or wellformed diagrams.

For c) overall (i.e. all three questions), see Fig. 13 for the results.

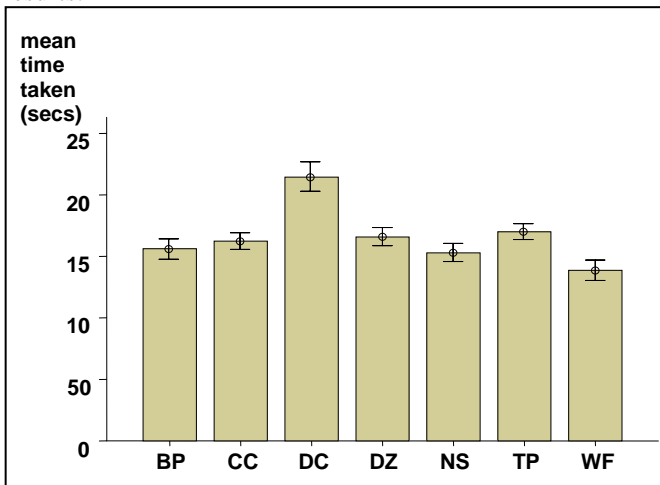


Fig. 13. Second study overall performance.

A Friedman non-parametric related-measures test produced a Chi-square value of 53.84. A post-hoc Nemenyi pair-wise comparison test revealed where the significant differences lie. Hence, in the case of all cases, we have evidence that:

- tasks on diagrams with duplicated curve labels take a longer time than tasks on diagrams with all the other properties;
- tasks on diagrams with either disconnected zones or triple points take a longer time than tasks on wellformed diagrams.

3.2.2 Preference Results

For the preference part of the study, we are interested the question “which properties do participants prefer?”. Participants were asked to assign a number between 1 and 7 for each diagram, with 1 as most preferred, 7 as least preferred. See Fig. 14 for the results.

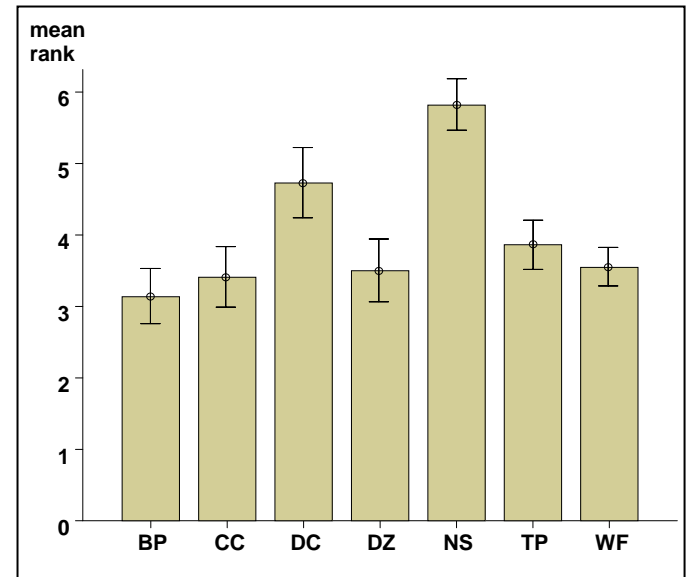


Fig. 14. Second study wellformedness property preference.

A Friedman non-parametric related-measures test produced a Chi-square value of 25.48. As this is greater than chi-square ($p=0.05, 6$) = 12.59, there are significant differences between the properties.

A post-hoc Nemenyi pair-wise comparison test revealed where the significant differences lie:

- the diagram with non-simple curves was least preferred over all others apart from the diagram with duplicated curve labels.

We then looked at the correlation of the two sets of data with the research question: “Is there a relationship between participants preference for a property and their performance using diagrams using that property?”, see Fig. 15 for the results.

The Pearson correlation between the time taken by a participant to answer a question using a diagram associated with one of the seven properties, and the preference rank (between 1 and 7) given to that property

by that participant is -0.006 ($p=0.809$), $N=1386$ – effectively no correlation at all. Thus: performance and preference are not correlated.

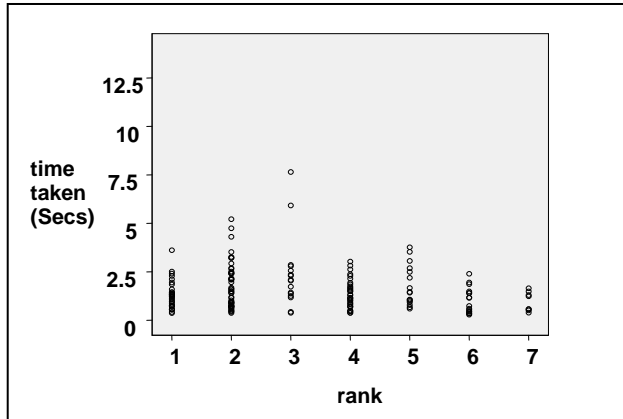


Fig. 15. Second study performance/preference correlation.

3.3 Second Study Analysis

The timing data indicates that there is a consistently poor performance with diagrams that have duplicated curve labels. In addition, disconnected zones perform worse than some other properties. These results indicate that there are problems when splitting an entity, whether it is a curve or zone, perhaps because of the increased cognitive load that results from forcing users to keep disjointed components in short-term memory.

There is evidence from the overall data that concurrency performs badly when compared to wellformed diagrams and brushing points, which does not appear when the data is broken down into the smaller targeted and non-targeted sets. It can be conjectured that concurrency makes lines harder to follow, as they merge into and out of concurrent sections. An alternative explanation is based on the shape of the curves in the concurrent diagram, which have to be more rectangular to ensure concurrency can be present, whilst drawing what we perceived to be an effective diagram. So perhaps ensuring that the concurrency was as non-disruptive as possible (as concurrency only occurred on straight line segments) affected other aspects of diagram comprehension.

As expected, wellformed diagrams perform consistently better than most properties. This is evidence to support the notion that wellformedness aids understanding of Euler diagrams. The exception is brushing points, for which there is no evidence that they disrupt the understanding of diagrams.

Looking at the classification into targeted and non-targeted data, a) and b), there is some consistency with duplicated curve labels performing badly. It is interesting that, even when the results of the question is not in the vicinity of the wellformedness property, the property still has an effect. Perhaps this is because the data around the property is still examined when the solution is being sought.

In terms of preference, the diagram with non-simple curves was least favoured over all others, apart from the

diagram with duplicated curve labels. Preference results did not correlate with the performance results. The reasons for lack of correlation may be down to participants' aesthetic perception, as other wellformedness properties may result in a superficially more attractive diagram over the relatively intricate diagram that results from non-simple curves.

3.4 Comparison Between the Two Studies

The notion that disconnected zones and concurrency are key wellformedness properties in user understanding of Euler diagrams is carried through both studies. This consistency implies that these two properties should be avoided where possible.

The initial study indicates that duplicated curve labels do not provide a barrier to understanding, however the second study found that there was a significant effect. This may be explained by the abstract data in the initial study, as opposed to concrete nature of the data in the second. These concrete tasks have a higher demand on memory as names were present in the diagram, rather than simple dots. This extra cognitive load may have made the extra requirement of identifying multiple curves more difficult. Another explanation could be the increased size of the diagrams in the second study. The disconnected components were then further apart, and so less easy to identify.

Both studies indicate that brushing points have no significant impact on interpreting Euler diagrams. In the initial study, brushing points were combined with other wellformedness properties, and so conclusions here are more difficult to make. However, the results in the second study, where the properties were independently changed, means that we can say that the presence of brushing points are unlikely to be an important consideration when designing diagrams for applications similar to that in the second study.

4 CONCLUSIONS AND FURTHER WORK

As with any controlled study, the generalizability of our results are constrained by our experimental parameters, in particular, the size of the diagrams, the nature of the participants, the questions asked, and the time limit given.

Despite these limitations, we have established some principles for the most effective presentation of Euler diagrams based on empirical evidence. Firstly, the second study indicates that wellformed diagrams increase the comprehension of the diagram, and so should be preferred where possible.

In terms of non-wellformed diagrams, the results from both studies suggest that it is harder to perform tasks on diagrams with either disconnected zones or concurrency compared to diagrams with other wellformedness properties broken, so these should be avoided where possible. The second study implies that tasks on diagrams with duplicated curve labels take longer than on diagrams with the other wellformedness properties broken, but this is not supported by the initial study, so this may be application dependent.

We note that the preference data indicates that diagrams that break the other wellformedness properties (except duplicated curve labels) are preferred by users over a diagram with non-simple curves. Hence, if user preference is most important, then non-simple curves should be avoided.

The results here have implications for the example at the start of this paper. Of the two versions of the same data in Fig. 1, Fig. 1b might be used from the perspective of performance as it has non-simple curves, rather than having concurrency as shown in Fig. 1a. However, if user preference is of most importance, then Fig. 1a might be the one that should be used.

There are also implications for automated Euler diagram generation. Several methods produce diagrams with significant amount of concurrency [4][16][21], so the results here might imply that alterations to the way visualizations are generated so as to minimize concurrency could improve users' understanding of the information.

This work is the first that gives an indication of the most appropriate use of wellformedness properties from the perspective of human comprehension in a concrete application area. More work is needed to compare the findings here with those for other applications and with current generation methods, as the specific usage of a diagram is likely to have an effect on how the wellformedness properties affect understanding. Rendering issues, such as the particular curve shapes used, could also be studied, to see if alternative kinds of layout could improve understanding over the diagrams we used in this study. Moreover, it may be that some of the results are a result of the curve shape that was used, particularly in the case of concurrency, which often has to be represented with less regular shapes (i.e. not circles or ellipses).

Other representation schemes have used rectilinear shapes [3] which consist of straight lines joined at right angles, and these might be compared with the smoother shapes used in this paper. The use of colour and shading is widespread when Euler diagrams are rendered, hence, the most effective rendering method with variations of these features could also be explored.

The lack of correlation between preference and performance in the second study raises interesting questions as to how to design diagrams, as what people like is not necessarily good for them. This introduces a dilemma for the designers of diagrams as preference is a major issue when communicating information, and it may be that a diagram that is preferred by users may be a better mechanism for information visualization than one that is actually better for interpreting data.

Finally, the large discrepancy between the results of the studies described here and work by other researchers [9] motivates further investigations to discover if the difference is in task, diagram design or some other factor.

ACKNOWLEDGMENT

We would like to thank the participants for assisting with this study. We would also like to thank Tomasz Manchura

for his work on the software we used in this study. This research was supported by the EPSRC under grant EP/E010393/1 "Visualization with Euler Diagrams". The project web page is <http://www.eulerdiagrams.com>. This work was also partially funded by the German Research Foundation (DFG) under grant GK-1042 "Explorative Analysis and Visualization of Large Information Spaces".

REFERENCES

- [1] J. Barwise, A. Shimojima. Surrogate Reasoning, *Cognitive Studies: Bulletin of Japanese Cognitive Science Society*, 4(2): 7-27. 1995.
- [2] F. Benoy, P. Rodgers. Evaluating the Comprehension of Euler Diagrams. In *11th International Conference on Information Visualization (IV 2007)*, pages 771-778. IEEE, July 2007.
- [3] S. Chow. Generating and Drawing Area-Proportional Venn and Euler Diagrams. PhD Thesis. University of Victoria, Canada. 2008.
- [4] S. Chow, F. Ruskey. Drawing area-proportional Venn and Euler diagrams. In *Proc. of Graph Drawing 2003*. LNCS 2912, pages 466-477. Springer-Verlag.
- [5] R. DeChiara, U. Erra, V. Scarano. VennFS: A Venn diagram file manager. In *7th International Conference on Information Visualization (IV 2003)*, pages 120-126. IEEE, 2003.
- [6] L. Euler. *Lettres à une Princesse d'Allemagne*, vol 2. 1761. Letters No. 102-108.
- [7] G. Farfel, W. Sousa. Repeat Victimization and Hot Spots: The Overlap and its Implication for Crime Control and Problem-Oriented Policing. *Crime Prevention Studies* vol. 12, pages 221-240, 2001.
- [8] A. Fish, B. Khazaei, C. Roast. Exploring Human Factors in Formal Diagram Usage. In: *Engineering Interactive Systems 2007*, LNCS 4940, pages 413-428.
- [9] A. Fish, B. Khazaei and C. Roast. User-comprehension of Euler Diagrams, *J. Visual Languages and Computing*, in press. doi:10.1016/j.jvlc.2011.01.002. 2011.
- [10] J. Flower and J. Howse. Generating Euler Diagrams, *Proc. Diagrams 2002*, LNAI 2317, Springer Verlag, pages 61-75.
- [11] S. Hughes. The Great British Venn Diagram. <http://qntm.org.uk>. Accessed 14 June 2011.
- [12] 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.
- [13] N. H. Riche, T. Dwyer. Untangling Euler Diagrams. *Infovis 2010, IEEE Trans. on Visualization and Computer Graphics* 16(6):1090-1099, 2010.
- [14] P. Rodgers, J. Howse, G. Stapleton, L. Zhang. Euler Graph Transformations for Euler Diagram Layout. *IEEE Symposium on Visual Languages and Human Centric Computing (VL/HCC 2010)*, pages 111-118, September 2010.
- [15] P. Rodgers, L. Zhang, A. Fish. General Euler Diagram Generation. In *Proc. Diagrams 2008*, LNCS 5223 pages 13-27.
- [16] P. Simonetto, D. Auber, and D. Archambault. Fully automatic visualisation of overlapping sets. *Computer Graphics Forum*, 28(3), 2009.
- [17] J. Soriano, K. Davis, B. Coleman, G. Visick, D. Mannino, N. Pride. The Proportional Venn Diagram of Obstructive Lung Disease: Two Approximations From the United States and the United Kingdom. *Chest* 124:474-481, 2003.
- [18] G. Stapleton, P. Rodgers, J. Howse, J. Taylor. Properties of Euler Diagrams. In *Proceedings of Layout of (Software) Engineering Diagrams*, Volume 7 of *Electronic Communications of the EASST*, 2007.
- [19] A. Treisman, J. Southera. Search Asymmetry: A Diagnostic for Preattentive Processing of Separable Features. *Journal of Experimental Psychology: General* 114(3):285-310. September 1985.
- [20] J. Venn (1880) On the Diagrammatic and Mechanical Representation of Propositions and Reasonings. *Phil. Mag.*(5)9, 1-18.
- [21] A. Verroust, M.-L. Viaud. Ensuring the drawability of Euler diagrams for up to eight sets. In *Proceedings of 3rd International Conference on the Theory and Application of Diagrams*, volume 2980 of LNAI, pages 128-141, Cambridge, UK, 2004. Springer.
- [22] C. Ziemkiewicz, R. Kosara. Beyond Bertin: Seeing the Forest despite the Trees. *IEEE Computer Graphics and Applications* September 2010, pages 7-11.

Peter Rodgers is a Senior Lecturer in the School of Computing at the University of Kent, UK. His research interests are in diagrammatic information visualization. He has published papers in topics such as automatic layout of diagrams, drawing schematic networks and novel visual languages. He is currently the recipient of a Royal Society Industrial Fellowship.

Leishi Zhang is a post-doctoral fellow in the Department of Computer and Information Science at the University of Konstanz, Germany. Her main research interests include visual analytics, data mining and graph visualization. She has published her research in a number of international journals and conferences relating to the area of data analysis and visualization.

Helen Purchase is a Senior Lecturer in the School of Computing Science at the University of Glasgow, UK. While her main interest is the evaluation of the visual presentation of graphs, she also involved in empirical research projects investigating a variety of visual stimuli.