Evaluating the Impact of Clutter in Linear Diagrams

Mohanad Alqadah[1], Gem Stapleton[1], John Howse[1], and Peter Chapman[2]

[1] University of Brighton, UK
{m.alqadah1, g.e.stapleton, john.howse}@brighton.ac.uk
[2] Edinburgh Napier University, UK
p.chapman@napier.ac.uk

Abstract. Linear diagrams are an effective way of visualizing sets and their relationships. Sets are visualized by a collection of straight line segments and the ways in which the lines overlap indicate subset and disjointness relationships. As with many visualization methods, linear diagrams can become cluttered. In previous research, we established a clutter measure for linear diagrams that was empirically shown to correlate with perceived clutter. The aim of this paper is to determine the impact of linear diagram clutter on user task performance. An empirical study was conducted with three levels of clutter. Surprisingly, we found that diagrams with a medium level of clutter had significantly slower task performance than low and high cluttered diagrams. Moreover, we found no significant performance difference between the low and high clutter. We concluded that clutter affects the interpretation of linear diagrams. A future research goal is to establish methods for controlling the level of clutter in linear diagrams, such as using multiple diagrams instead of a single diagram, when visualizing sets.

Keywords: linear diagrams, clutter, diagram comprehension

1 Introduction

Presenting a lot of information in a diagram and the layout choices made when drawing a diagram can result in visual clutter, known to sometimes be a barrier to cognition. It is therefore essential to understand clutter and how to reduce clutter when needed. Ellis and Dix state that it is "important to have a clear understanding of clutter reduction techniques in order to design visualizations that can effectively uncover patterns and trends" [7]. Bertini *et al.* have a similar standpoint, devising an approach to analyze and reduce clutter in an information visualization context [3].

Visualizing information using diagrams can have huge benefits over textual notations, provided the diagrams are effective [12]. In the context of visualizing sets, there are several methods that can be used, including Euler, Venn, and linear diagrams, with the first two being well-known. In the context of Euler and Venn diagrams, clutter was shown to have an impact on user comprehension [1]. Linear diagrams were introduced by Leibniz in 1686 [6] and they are
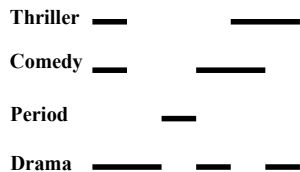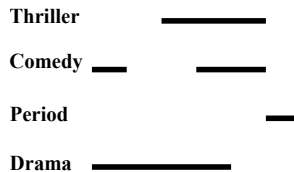
Fig. 1: A linear diagram.

Fig. 2: Altering overlap order.

similar to parallel bargrams [20] and double decker plots [9]. Recently, Chapman *et al.* [5] empirically established that linear diagrams more effectively support task performance than Euler and Venn diagrams. Sato and Mineshima [18] have also shown that linear diagrams are superior to the linguistic representations of syllogisms in the context of logical reasoning. These results motivate the need to understand the role of clutter in linear diagrams to better place us to exploit their cognitive advantages.

A linear diagram consists of parallel line segments drawn horizontally. Each set presented in the diagram is represented by the line segments that share their y-coordinate. For example, Fig. 1 represents four sets using eight line segments. The set Thriller is represented by two line segments. Line segments for different sets can occupy the same vertical space, known as an *overlap*, to represent the intersection between the sets. For example, in Fig. 1 the left-most overlap contains line segments for the sets Thriller, Comedy, and Drama but not Period. If two sets do not share any overlaps then these two sets are disjoint. For instance, in Fig. 1 set Period is disjoint from all the other sets in the diagram. Moreover, if all of the line segments for a set occur in overlaps with line segments from another set then the former is a subset of the latter.

Research has suggested that to use linear diagrams successfully, we need to understand when they can be effectively interpreted [17]. In this paper we set out to establish the impact of linear diagram clutter on user task performance, in term of time and accuracy. The structure of the rest of this paper is as follows. A theoretical measure of clutter in linear diagrams is given in section 2. We present the experiment design in section 3. Section 4 describes how we executed the experiment and section 5 presents the results of the study. In section 6, we address the threats to validity. In section 7, we offer conclusions and discussion on future work. The diagrams used in the study, along with the raw data collected, can be found at `https://sites.google.com/site/msapro/phdstudyfour`.

2   Clutter in Linear Diagrams

In previous work we defined a clutter measure, called the *line score*, for linear diagrams: each set contributes n to the line score, where n is the number of line segments that are used to represent that set [2]. We established that the line score clutter measure meets user perceptions: diagrams with a higher line score were generally perceived as more cluttered.
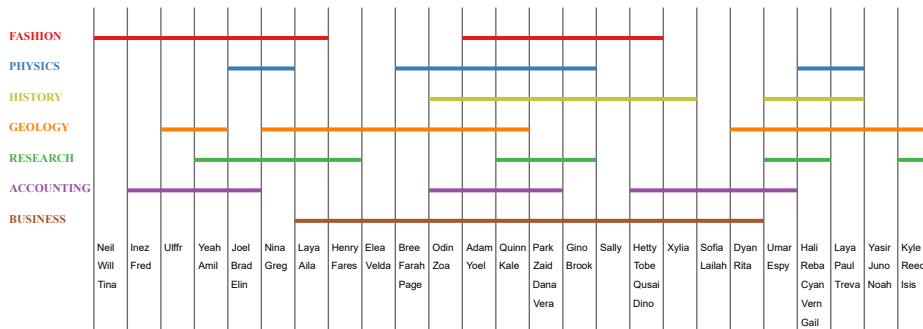
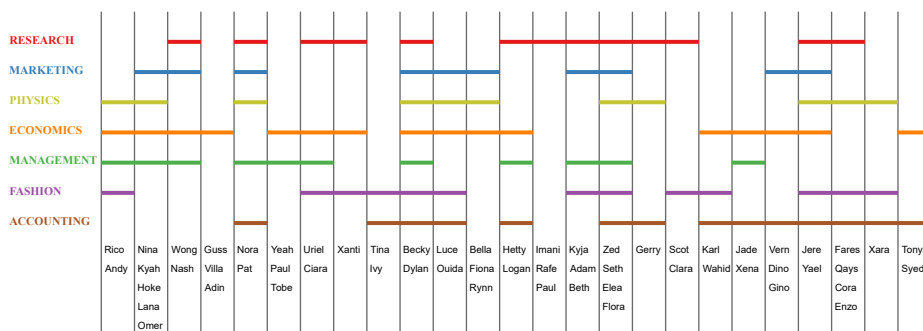Fig. 3: Low cluttered linear diagram with 7 sets.

Fig. 4: Medium cluttered linear diagram with 7 sets.

For example, the diagram in Fig. 1 has a line score of 8; the contributions of each set to the score are 2, 2, 1, and 3, from top to bottom. We can change the order of the overlaps to get the semantically equivalent diagram in Fig. 2. This change in the order of overlaps reduces the clutter score to 5. As a more complex example, the diagrams in Figs 3 to 5 exhibit three different line scores, with Fig. 3 being the least cluttered and Fig. 5 being most cluttered. In general, we can permute the order of the overlaps without altering the information being represented but leading to different clutter scores. Therefore, if we understand the impact of clutter on task performance, we can chose between competing overlap orders in order to best support task performance.

3 Experiment Design

To answer the research question 'does linear diagram clutter affect user comprehension?' we designed a within-group empirical study, building on an established measure of clutter, the line score, for linear diagrams [2]. Each participant was asked to answer questions about the information in a set of linear diagrams which had varying levels of clutter. Consistent with other research contributions towards understanding user comprehension, such as [4, 11, 13, 15], we used

Fig. 5: High cluttered linear diagram with 7 sets.

research software that was designed to record user performance data while performing the experiment. Two primary variables were recorded: the time taken to answer the question and whether the question had been answered correctly. We set a time limit of two minutes to answer each question and all timeouts were recorded. If clutter level affects user comprehension of linear diagrams then we would expect to see a significant difference between the means of the time taken to answer the questions or the error rates over the different levels of clutter.

3.1 Study Details and Question Types

The participants were staff or students from the School of Computing, Engineering and Mathematics at the University of Brighton; none of them were members of the authors' research group. Consistent with Rodgers *et al.* [16], the diagrams were drawn with pseudo-real data, conveying information about students' interests in university subjects. For example in Figs 3, 4 and 5, the line segments for each set represent a subject. The student names are written under the overlap corresponding to the subjects in which they are interested. This design choice was to ensure the context of the information presented was familiar to the participants, removing any learning barrier that could arise through participants learning a new context. The data presented in the diagrams was hypothetical to ensure that prior knowledge cannot be used.

Subject names consisted of one word, did not sound similar to each other and did not share the same first letter, because that could confuse the participants as we found in our previous research [1]. Furthermore, each subject name was placed on the left side of the diagram and aligned with, and took the same colour as, the associated line segments. The order of the subject names was random and no two diagrams had the same order. This was important to make sure the participants had to read the question and the diagram each time, reducing learning effect that could impact on performance.

The student names that were presented in the diagrams were a randomized mixture of both male and female names across a variety of ethnicities, to reduce any bias that could occur if we chose the names manually. Moreover, the names
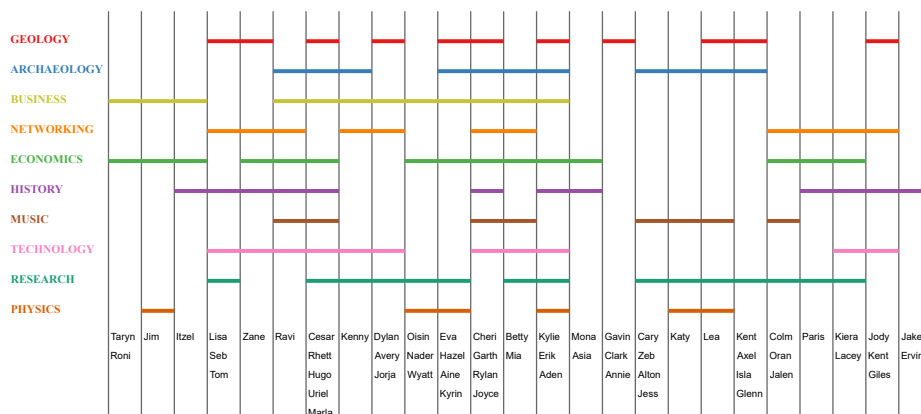
Fig. 6: Medium cluttered linear diagram with 10 sets.

were first name only and of a similar length (three, four or five letters) to reduce the risk of having a name that was more prominent than the others (e.g very long names). We avoided using names more than once in the same diagram. Moreover, the names that appear in the same overlap sounded different and started with different first letters. This design choice was made to limit any confusion and make the names easier to remember.

Three types of multiple choice questions were used in the study: 'Who', 'How many', and 'Which', consistent with previous research [1, 4, 16] [3]. There were five choices of answer of which one was correct. Example questions are:

1. **Who** of the following is taking PHYSICS, HISTORY, GEOLOGY, RE-SEARCH and nothing else?
2. **How many** students are taking GEOLOGY, ARCHAEOLOGY, MUSIC, RESEARCH, PHYSICS and nothing else?
3. **Which** one of the following modules is being taken by 23 students?

Questions 1, 2, and 3 were asked of the diagrams in Figs 3, 6, and 7 respectively. The set names that appear in the 'Who' and 'How many' questions, as well as in the answers, match the same order in the relevant diagram from top to bottom. This choice should make it easier for the participants to find the required sets to answer the questions. After choosing an answer for the 'How many' questions, the participants were asked to type the names of the students they had counted. For their answer to be classed as correct, the names typed in had to be correct. For 'Which' questions, there was only one set in the diagram that had the number of students' names specified in the question.

---

[3] The empirical studies on linear diagrams in [17] used tasks that required people to identify subset, disjointness, and non-empty intersection relationships between sets and are, thus, different from the tasks we use in this study.
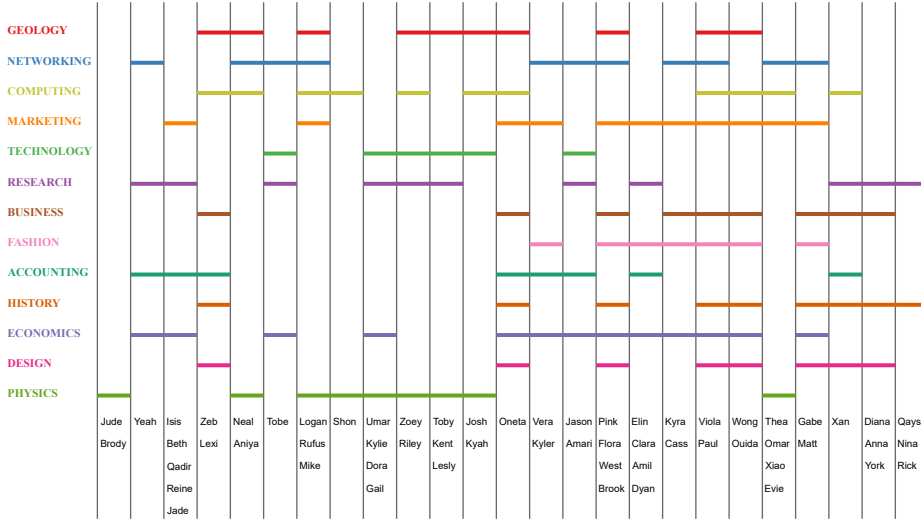
Fig. 7: High cluttered linear diagram with 13 sets.

3.2   Levels of Diagram Clutter

We used three different levels of clutter in the study, as measured by the line score, to establish whether *diagram clutter* impacts user performance when interpreting linear diagrams. It was felt important that the diagrams were sufficiently complex to demand notable cognitive effort by the study's participants. Given this, the diagrams were chosen so that their line scores (LSs) were: $10 - 20$ (low LS), $30 - 40$ (medium LS), and $50 - 60$ (high LS). For example, the three diagrams in Figs 3, 6 and 7 have LSs of 18, 40, and 60 respectively. For the study, 27 diagrams were drawn (nine for each level of clutter). These 27 diagrams were created from nine initial diagrams, which were then redrawn with different overlap orders to create different levels of clutter.

3.3   Clutter by Number of Sets

To widen the generality of the results from our study, the number of sets visualized in the diagrams varied: they had either 7, 10 or 13 sets. The distribution of the number of sets to clutter scores is shown in table 1. For example, the diagrams in Fig. 3, 6, and 7 represent 7, 10, and 13 sets respectively. The three types of questions were asked from each set of diagrams. For instance, as can be seen in table 1, the three types of questions were asked of the information in the diagrams drawn with seven sets (d1, d2, and d3).

By including a variety of numbers of sets represented, we will be able to determine whether there is an interaction with the clutter score. Suppose, for example, that we have two diagrams, d_7 and d_{10}, each with a clutter score of 20, where d_7 and d_{10} represent 7 and 10 sets respectively. Then d_7 has, on average 2.9

Table 1: The characteristics of the diagrams.

Diagram reference	d1	d2	d3	d4	d5	d6	d7	d8	d9
Question type	Who	How many	Which	Who	How many	Which	Who	How many	Which
Number of sets	7 Sets			10 Sets			13 Sets		
Low clutter	$10 - 20$ line score								
Medium clutter	$30 - 40$ line score								
High clutter	$50 - 60$ line score								

lines per set, whereas d_{10} has just 2 lines per set. In general, when the diagram clutter is high and the number of sets is small, the diagram has a high number of line segments in a small (vertical) space, as measured by the number of sets represented. When the number of sets increases there is more vertical space to fit the same number of line segments. Our study allows us to explore whether the measure of clutter relative to the number of sets influences cognition.

3.4 Linear Diagram Layout and Characteristics

As can be seen in table 1, the diagrams were divided into nine characteristic types (d1 to d9); the number of sets represented appears under each characteristic type. For each characteristic type there were three diagrams: a diagram for each diagram clutter level. The three diagrams of each characteristic type were allocated the same type of question. So, for example, d1 represented 7 sets, and corresponded to a 'Who' question, and was drawn in three different ways, to give low, medium and high levels of clutter. The number of data items (student names) was chosen to be large enough so that answering the question required cognitive effort. There were one to five students names within each overlap and all the diagrams had 60 names. We randomly distributed the data items within the overlaps in the diagrams. However, the number of items in each overlap was the same in the three diagrams of each characteristic type. For example, the low clutter diagram for d1 may have an overlap with, say, three student names. In the medium clutter diagram for d1, this overlap need not occupying the same vertical space but will still have three names written under it.

We were careful to control the layout features for the diagrams drawn for the study. Therefore, we adopted the following six layout guidelines to ensure consistent layout features that reflected best-practice in linear diagram drawing [17]:

1. The diagrams were drawn in a horizontal direction.
2. All diagrams were drawn with coloured straight line segments and no two sets appearing in the same diagram had the same colour.
3. The sets in the diagrams had the same distances between them.
4. The overlaps in the diagrams had the same length, fixed at 50 pixels.
5. Vertical grid lines with grey colour' were used to represent the beginning and the end of each overlap.
6. The line segments height was set to 6 pixels.

Guidelines 1-6 are from [17], but they are not sufficient to control variability across diagrams in our study. We stipulate five more layout guidelines:

7. The order of the colour of sets is the same in all the diagrams.
8. All the diagrams had 25 overlaps.
9. The stroke width for the vertical grid lines was set to 2 pixels.
10. The set labels were written using upper case letters in Time New Roman, 14 point size, font in bold.
11. Data items were presented using lowercase letters, except that the first letter was capitalized, with black colour and Ariel 12 point size font.

Consistent with [5, 8], a palette of thirteen colours was generated using color-brewer2.org (accessed January 2016). Colour generation using the Brewer colour palette is recognized as a valid approach for empirical studies, such as in the context of maps [19]. So that the colours were distinguishable, but not sequential or suggestive (e.g. increasingly vivid shades of red used to denote heat), they were generated using the 'qualitative' option, based on work by Ihaka [10].

4   Experiment Execution

The initial pilot study was conducted with five participants (4 M, 1 F, ages $18 - 38$). The pilot study showed that the two minute time limit to answer each question was enough for most participants, only one participant had 3 timeouts and 1 error, and another participant had 5 errors. The three remaining participants answered all the questions correctly in the time limit. However, a problem was revealed with the number of items that appear under the overlaps in the diagrams: they were not equally distributed between the three diagrams for each characteristic type. This unevenness may have caused bias across the clutter levels, because the speed or accurately of answering questions could depend on the distribution of items. We fixed this issue by equalizing the number of items in the same overlaps across the three diagrams for each characteristic type. For example, in the left-most overlap in the diagram in Fig. 3 there are three students names (Neil, Will, and Tina), the equal overlap in the middle part of Fig. 4 contains three students names (Imani, Rafe, and Paul) as well, and in same overlap in Fig. 5 also has three names (Glen, Fred, and Owen). The main experiment was carried out with 30 participants (19 M, 11 F, ages $17-34$). Each participant was given £6, in the form of a canteen voucher, to take part.

There were three phases for the experiment. First is the introduction and the paper-based training phase, the experimental facilitator used the introduction script to introduce the participant to linear diagrams and the type of questions they will see during the experiment. The participants then started the second, computer-based training, phase where the participants had the chance to use the experiment software to answer three questions in the two minutes given for each question without interruption from the experiment facilitator. If a training question was answered incorrectly, the experimental facilitator went through the question with the participant. If the participant wished to continue then we

proceeded to the third (main) study phase where we collected the quantitative data for our analysis. Each participant signed a consent form to say they would like to participate in the study. The participant was not interrupted or given any help during this phase. After answering all the 27 questions, the participant was given the debrief script, including the contact information of the experimental facilitator so that they could access the results of the study.

5   Experiment Results

The results are based on the data collected from the 30 participants in the main study phase, each answering 27 questions, yielding $27 \times 30 = 810$ observations.

5.1   Overall Impact of Clutter

We analyzed the time and error data to determine whether significant differences exist, overall, between the three clutter levels. This analysis allowed us to determine whether clutter level significantly impacts on user task performance.

Analysis of Time Data The observed means are 42.93 (sd: 27.65) seconds for low diagram clutter, 46.35 (sd: 25.89) seconds for medium diagram clutter, and 42.23 (sd: 24.94) seconds for high diagram clutter. The data are not normally distributed and have a high skew. To make the data suitable for conducting an ANOVA, it was subjected to a log transformation (resulting skewness: -0.05). Therefore, the ANOVA was conducted in this transformed data, yielding a p-value of 0.002 when testing for differences between the means for diagram clutter. We therefore conclude that the mean time taken to interpret linear diagrams alters significantly as the clutter score changes.

To reveal how the significant differences arose, we conducted a Tukey test to rank the clutter levels. Surprisingly, this test revealed that the difference between the mean time taken to interpret linear diagrams with low diagram clutter versus high diagram clutter *is not* significant. However, the difference between the mean time taken to interpret linear diagrams with medium diagram clutter versus low and high diagram clutter *is* significant. In conclusion, medium diagram clutter has a significantly higher mean time as compared to low and high clutter. The medium clutter level questions took roughly 3.5 to 4 seconds longer, on average, to answer than the low and high clutter questions.

Analysis of Errors Of the 810 observations (30 participants $\times$ 27 questions) there were a total of 31 errors and 18 timeouts which gives an error rate of 3.83% and a timeout rate of 2.22%. Timeouts were counted as neither correct responses nor errors. Table 2 shows the observed number of correct answers, errors and timeouts for each clutter level. We performed a χ^2 test, giving a p-value of 0.287, which is not significant [4]. We conclude that diagram clutter does not affect the comprehension of linear diagrams in terms of accuracy.

---

[4] Treating timeouts as errors changes the p-value to 0.266.

Table 2: The observed correct answers, errors and timeouts for each clutter level.

Clutter Level	Correct	Errors	Timeouts	Total
Low Clutter	249	11	10	270
Medium Clutter	254	12	4	270
High Clutter	258	8	4	270
Total	761	31	18	810

5.2 Interaction between Clutter and Number of Sets

Analysis of Time As discussed earlier, by including a variety of numbers of sets we are able to determine whether there is an interaction between the number of sets and the clutter score. The following analysis seeks to address whether the measure of clutter relative to the number of sets influences cognition. The mean times for each clutter level (along with the standard deviations) are shown in Table 3. It can be seen that, for 10 and 13 sets, the medium level mean time is higher than that for the low and high clutter levels, consistent with our overall analysis. However, in the case of 7 sets, the mean times are all similar. This may indicate that the level of cognitive difficult associated with the number of sets represented in relation to the level of clutter is low in this case, and perhaps suggests that clutter has more of an effect as the number of sets rises.

Table 3: Time data by clutter level and number of sets.

	Number of Sets		
Clutter Level	7	10	13
Low	49.37 (30.24)	41.36 (25.29)	38.06 (26.22)
Medium	47.67 (29.66)	48.20 (25.11)	43.19 (22.37)
High	46.39 (27.73)	39.81 (25.83)	40.49 (20.42)

Our ANOVA revealed a significant interaction, with $p = 0.001$. In table 4, the statistical results of the pairwise comparisons of the interaction between diagram clutter and number of sets. This reveals that:

1. Participants took significantly longer to answer questions of the 10 set diagrams with medium clutter than the 10 and 13 set diagrams with high and low clutter respectively.
2. Participants took significantly longer to answer questions of the 7 set diagrams with low and medium clutter than the 10 and 13 set diagrams with high and low clutter respectively.
3. Participants took significantly longer to answer questions of the 7 set diagrams with high clutter than the 13 set diagrams with low clutter.

The results suggest that the combination of diagram clutter and the number of sets impacts on the time taken to interpret linear diagrams.

Analysis of Errors We will compare the error rates between the number of sets represented. This will reveal whether increasing the number of sets, whilst keeping the clutter level constant, impacts on cognition. However, we note that the following analysis should be treated with caution, as the observed values in some cases are very low. Firstly, in the case of low clutter, conducting a χ^2 test yields a p-value of 0.819, with the data presented in table 5. We conclude that there is no significant difference between performance, in terms of errors, as the number of sets varies when the clutter level is low. However, we saw in our interaction analysis of the time data that low cluttered diagrams with 7 sets took significantly longer for participants than the 10 and 13 set diagrams.

In the case of medium clutter, conducting a χ^2 test yields a p-value of 0.068, with the data presented in table 5. We again conclude that there is no significant difference between performance, in terms of errors, as the number of sets varies. This is consistent with our time analysis, where we found no significant performance difference between the medium clutter diagrams as the number of sets varied (see table 4).

Lastly, for high clutter, conducting a χ^2 test yields a p-value of 0.823, with the data presented in table 5. We again conclude that there is no significant difference between performance, in terms of errors, as the number of sets varies. This is again consistent with our time analysis, where we found no significant performance difference between the high clutter diagrams as the number of sets varied (see table 4).

5.3   Discussion

We now seek to explain the findings of the study. We found no impact of clutter on accuracy. The low error rates could be partially attributed to our design choices: the diagrams were drawn based on guidelines that have been empirically established to have positive impact on user comprehension [17].

The only statistically significant differences found arose through our analysis of the time data, which revealed, surprisingly, that a medium level of clutter generally led to poor task performance. To understand why this might have

Table 4: Grouping information using the Tukey method and 95.0% Confidence.

Clutter Level	No. of Sets	N	Mean (Log Time)	Mean (Seconds)	Grouping
Medium	10	90	1.627	48.20	A
Low	7	90	1.601	49.37	A
Medium	7	90	1.596	47.67	A
High	7	90	1.585	46.39	A B
Medium	13	90	1.571	43.19	A B C
High	13	90	1.559	40.49	A B C
Low	10	90	1.550	41.36	A B C
High	10	90	1.513	39.81	B C
Low	13	90	1.493	38.06	C

Table 5: Observed values for low clutter level by number of sets.

Number of Sets	Correct	Errors	Timeouts	Total
7	82	4	4	90
10	82	5	3	90
13	85	2	3	90
Total	249	11	10	270

Table 6: Observed values for medium clutter level by number of sets.

Number of Sets	Correct	Errors	Timeouts	Total
7	81	5	4	90
10	87	3	0	90
13	86	4	0	90
Total	254	12	4	270

arisen, we begin by making some observations about the relationship between clutter, the number of sets represented, and the lengths of the corresponding line segments. Now, the average number of line segments per set is $\frac{clutter\ score}{number\ of\ sets}$. Therefore, as the clutter score increases, the length of the line segments used decreases. As we used a fixed number (25) of overlaps this means that diagrams with a high clutter score will exhibit more short lines, relatively speaking, than diagrams with a low clutter score. Diagrams with a medium level of clutter will fall somewhere in the middle. Consequently, in low clutter diagrams, many sets are typically represented by just one or two line segments. Medium clutter diagrams represent sets by around 2 to 3 line segments and High clutter diagrams represent sets by around 3 to 5 line segments. Therefore, we might expect many long line segments in the low clutter diagrams and short line segments in the high clutter diagrams. However, the medium clutter diagrams having will have more of a mix of the two, and therefore perhaps appear less uniform in layout. This can be seen in Fig. 3 (low clutter: many long line segments), Fig. 5 (high clutter: many short line segments) and Fig. 4 (medium clutter: more variability in line segment length).

Through our statistical analysis, we have demonstrated that extracting information takes significantly longer from a medium cluttered diagram than from a low or high cluttered diagram. We can infer, then, that understanding linear diagrams with a mixture of short and long line segments is more difficult than understanding diagrams which predominantly use only long, or short, line segments. We also found that there is no significant difference between the high cluttered diagrams and the low cluttered diagrams, overall. We hypothesize that different reading strategies, which we call *reading vertically* and *reading horizontally*, are employed for each type of diagram. The short line segments in highly cluttered diagrams could make it easier to identify an overlap by reading vertically, since the short line segments produce visually distinct blocks. We can thus easily associate an overlap with its items. By contrast, when a diagram

Table 7: Observed values for high clutter level by number of sets.

Number of Sets	Correct	Errors	Timeouts	Total
7	87	2	1	90
10	85	4	1	90
13	86	2	2	90
Total	258	8	4	270

is low cluttered it consists of long line segments which could make it easier to read horizontally, allowing an easy overview of the entire set. It might be that a mixture of these strategies best helps read medium cluttered diagrams, or that neither are ideal, increasing cognitive load. Further work is needed to establish whether people actually employ these reading strategies generally.

6 Threats to Validity

We now discuss the limitations of the study and how threats were managed to minimize their impact on the results. We categorized the threats to validity as internal, construct, and external as in [14]. In regard to the internal validity, the following factors have been considered during the study design:

1. *Motivation effect*: this could a threat if the participants did not freely volunteered to take part. To manage this effect, the participants were invited to take part as volunteers and they were given £6 refreshment voucher for their time.
2. *Laboratory*: it is important to expose all participants to the same environment in this type of study. The experiment took place in an usability laboratory where the environment was free from noise and interruption and all participants were treated in the same way.
3. *Fatigue effect*: fatigue could be a threat if the participants where exposed to repeated tasks for a long time. To control this, the study was designed to finish in less than one hour.
4. *Questions order*: we presented the questions in a random order for each participant to reduce any effect (e.g. tiredness, learning) owing to order.
5. *Layout variation*: many choices about diagram layout were made, such as the colour and the thickness of the lines [17]. To manage the threat of layout variations, we followed drawing guidelines.

Construct validity considers whether the identified dependent (primary) and independent (secondary) variables are the correct choices. Consistent with other researchers who studied user comprehension, for instance [4, 11, 13, 15], comprehension is measured by the time taken to answer the questions and the number of errors made. In addition, we also considered *false negatives*: these could arise if a participant selected the wrong answer while reading it to be the correct answer. To mange this, the similarity of university subject (i.e. set) names and student

names was controlled. Furthermore, when participants were asked a 'How many' question, they were required to type in the names of the students they had counted to reduce *false positives*.

A secondary variable is the choice of the diagrams. It could be a threat if participants did not spend enough time understanding the diagrams before answering the questions. To manage this threat we created diversity in the diagrams and designed them to convey an appropriate level of information in order to require cognitive effort when answering questions.

With regard to external validity, the following factor indicate the limitations of the results and to which extent we can generalize them. It is a common threat to the external validity of user studies if the participant have not represent a wider population. All participants were staff or students from the University of Brighton with little or no previous experience of using linear diagrams. The question styles could be a threat if there was no variety in the questions asked of the linear diagrams. To allow more generalization to the study we used three styles of questions.

7   Conclusion

In this paper, we determined whether, and how, clutter affects user understanding when interpreting linear diagrams. The results show that clutter does affect the comprehension of linear diagrams. Surprisingly, increasing clutter does not always negatively affect comprehension: the medium cluttered diagrams required significantly longer to interrogate than low or high cluttered diagrams. Moreover, we found no significant difference between the time taken to interpret high or low cluttered diagrams. Furthermore, the error data analysis showed that there is no significant relationship between error rates and complexity of the diagram. In summary, we found that diagram clutter affects time taken to interpret information in linear diagrams, but it does not affect the accuracy with which participants answered questions about the data visualized in linear diagrams.

These results inform the direction of further research into the effect of the complexity of linear diagrams on comprehension. For instance, it would be interesting to establish strategies people employed for reading linear diagrams, as hypothesized in our discussions above. A further fruitful step for this research is to determine whether it is beneficial to use multiple linear diagrams to convey information, rather than a single diagram, to control the amount of clutter present in any individual diagram, informed by the results of this study.

References

1. Alqadah, M., Stapleton, G., Howse, J., Chapman, P.: Evaluating the impact of clutter in Euler diagrams. In: Diagrammatic Representation and Inference, pp. 108–122. No. 8578 in LNAI, Springer (2014)
2. Alqadah, M., Stapleton, G., Howse, J., Chapman, P.: The perception of clutter in linear diagrams. In: Diagrammatic Representation and Inference (accepted for). LNAI, Springer (2016)

3. Bertini, E., Dell'Aquila, L., Santucci, G.: Reducing infovis cluttering through non uniform sampling, displacement, and user perception. In: Electronic Imaging 2006. pp. 60600L–60600L. International Society for Optics and Photonics (2006)
4. Blake, A., Stapleton, G., Rodgers, P., Cheek, L., Howse, J.: Does the orientation of an Euler diagram affect user comprehension? In: The 18th International Conference on Distributed Multimedia Systems, International Workshop on Visual Languages and Computing (2012)
5. Chapman, P., Stapleton, G., Rodgers, P., Micallef, L., Blake, A.: Visualizing sets: an empirical comparison of diagram types. In: Diagrammatic Representation and Inference, pp. 146–160. No. 8578 in LNAI, Springer (2014)
6. Couturat, L.: Opuscules et fragments inédits de Leibniz. Felix Alcan (1903)
7. Ellis, G., Dix, A.: A taxonomy of clutter reduction for information visualisation. IEEE Transactions on Visualization and Computer Graphics 13(6), 1216–1223 (2007)
8. Harrower, M., Brewer, C.: Colorbrewer.org: An online tool for selecting colour schemes for maps. Cartographic Journal 40(1), 27–37 (2003)
9. Hofmann, H., Siebes, A., Wilhelm, A.: Visualizing Association Rules with Interactive Mosaic Plots. In: 6th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 227–235. ACM (2000)
10. Ihaka, R.: Colour for presentation graphics. In: In 3rd Int. Workshop on Distributed Statistical Computing (2003)
11. Isenberg, P., Bezerianos, A., Dragicevic, P., Fekete, J.: A study on dual-scale data charts. In: IEEE Transactions on Visualization and Computer Graphics. pp. 2469 – 2478. IEEE (2011)
12. Larkin, J., Simon, H.: Why a diagram is (sometimes) worth ten thousand words. Journal of Cognitive Science 11, 65–99 (1987)
13. Puchase, H.: Which aesthetic has the greatest effect on human understanding? In: 5th International Symposium on Graph Drawing. pp. 248–261. Springer (1997)
14. Purchase, H.C.: Experimental human-computer interaction: a practical guide with visual examples. Cambridge University Press (2012)
15. Riche, N., Dwyer, T.: Untangling Euler diagrams. IEEE Transactions on Visualization and Computer Graphics 16(6), 1090–1099 (2010)
16. Rodgers, P., Zhang, L., Purchase, H.: Wellformedness properties in Euler diagrams: Which should be used? IEEE Transactions on Visualization and Computer Graphics 18(7), 1089–1100 (2012)
17. Rodgers, P., Stapleton, G., Chapman, P.: Visualizing sets with linear diagrams. ACM Transactions on Computer-Human Interaction (TOCHI) 22(6), 27 (2015)
18. Sato, Y., Mineshima, K.: How diagrams can support syllogistic reasoning: An experimental study. Journal of Logic, Language and Information 24(4), 409–455 (2015)
19. Silva, S., Madeira, J., Santos, B.: There is more to color scales than meets the eye: A review on the use of color in visualization. In: 11th International Conference on Information Visualization. pp. 943–950. IEEE Computer Society (2007)
20. Wittenburg, K., Lanning, T., Heinrichs, M., Stanton, M.: Parallel Bargrams for Consumer-based Information Exploration and Choice. In: 14th annual ACM symposium on User interface software and technology. pp. 51–60. ACM (2001)