The Perception 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 representing sets and their relationships. The topological and graphical properties of linear diagrams can affect perceived relative levels of clutter. This paper defines four different measures of clutter for linear diagrams. Participants in an empirical study were asked to rank linear diagrams according to their perception of clutter. We analyzed the correlation between how the clutter measures ranked linear diagrams compared to the overall ranking derived from the participants' perceptions. We concluded that the clutter measure which counts the number of line segments best matches participants' perception.

Keywords: Linear diagrams, clutter, diagram comprehension

1 Introduction

Representing information using diagrams can have huge benefits, but only if the diagrams themselves are effective. One aspect of the effectiveness of diagrammatic communication is related to clutter. If diagrams appear cluttered then their visual appeal and ability to support end-users with understanding the represented information can be reduced. Hence, there is clearly a need to theoretically understand what it means for diagrams to be cluttered and the impact of clutter on task performance. This paper is concerned with clutter in linear diagrams, recently shown to be superior to Euler and Venn diagrams when users perform set-theoretic tasks [1] and to linguistic representations of syllogisms [6].

A linear diagram consists of horizontal line segments drawn parallel to the x-axis. Each set is represented by the line segments that share their y-coordinate. For example, Fig. 1 represents five sets using six line segments. Line segments for

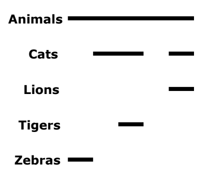

Fig. 1: A linear diagram.

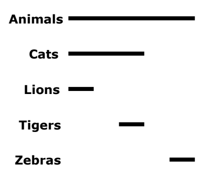

Fig. 2: Altering overlap order.

different sets can occupy the same vertical space, known as an *overlap*. Where an overlap contains line segments for sets $A_1, \ldots, A_n$ and does not contain line segments for $B_1, \ldots, B_m$, the overlap represents the information that $A_1 \cap \ldots \cap A_n \cap \overline{B_1} \cap \ldots \cap \overline{B_m}$ is non-empty. Moreover, if a set intersection is not represented by an overlap then that set intersection is empty. For example, in Fig. 1 the left-most overlap contains line segments for the sets Animals and Zebras but not Cats, Lions or Tigers. Thus, $\text{Animals} \cap \text{Zebras} \cap \overline{\text{Cats}} \cap \overline{\text{Lions}} \cap \overline{\text{Tigers}} \neq \emptyset$.

There is currently no understanding of what constitutes a cluttered linear diagram. This paper sets out to address the question of how to measure perceived clutter in linear diagrams. We introduce four measures of clutter in section 2. The design of our experiment to determine whether any of these four measures correlate with users' perception of clutter is described in section 3. We present the analysis and results in section 4 and conclude in section 5. The diagrams used in the study, along with the raw data collected, can be found at `https://sites.google.com/site/msapro/phdstudythree`.

2 Four Measures of Clutter for Linear Diagrams

Our first measure, called the *contour score* (CS), is based on a measure of clutter established when counting zones in Euler diagrams [2]. We can easily adapt this measure to linear diagrams as overlaps in linear diagrams directly correspond to zones in Euler diagrams. The CS for linear diagrams is computed as follows: each overlap contributes n to the contour score, where n is the number of lines in the overlap. For example, each diagram in Fig. 3 has six overlaps, identified by the use of grid lines, and have a CS of 11. In these four diagrams, the overlaps are annotated with their contribution to the contour score under the diagram, indicated by the CS label. These diagrams represent *the same information* as each other and have *the same CS*. However, they are syntactically different and, so, there may be differences in how people perceive their relative levels of clutter.

Our first new measure of clutter specifically designed for linear diagrams is the *line score* (LS): each set contributes n to the line score, where n is the number of line segments that are used to represent that set. For example, in Fig. 3, the diagrams (v1) and (v2) both have a LS of 8; the sets are annotated with their contribution to the LS under the column labelled LS. However, the

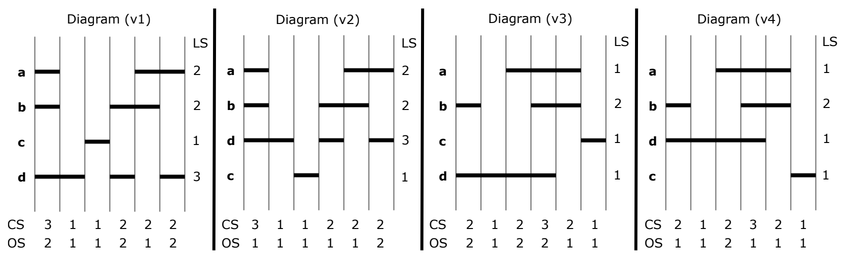

Fig. 3: Different measures of clutter in linear diagrams.

other two diagrams both have a LS of 5; the LS is lower because the overlaps are in a different order, leading to fewer line breaks.

While LS can be altered by changing the order of the overlaps, we also have the choice of reordering the horizontal lines in the diagram. Such reordering changes the visual appearance of the diagram and could also be linked to perceived clutter. Therefore, our third new measure of clutter, called the *overlap score* (OS), captures the 'vertical clutter' in linear diagrams; LS can be thought of as capturing the 'horizontal clutter'. The OS of linear diagrams can be computed from the blocks of lines in the overlaps: each overlap contributes n to the overlap score, where n is the number of blocks of lines in the overlap. For example, in Fig. 3, the diagrams (v1) and (v3) both have an OS of 9. The first overlap in (v1) has two blocks of lines: the two lines for sets 'a' and 'b' are a block of lines, then there is no line for the set 'c', and finally a further block, comprising a single line, for the set 'd'. Each overlap is annotated with its contribution to the overlap score. The diagrams (v2) and (v4) both have an OS of 7. The overlap score is lower than (v1) and (v3) because of the different orders of the represented sets.

Our last clutter measure is the combined score of the LS and the OS of linear diagrams, which we call the *line-and-overlap score* (LOS). This clutter measure is designed to capture both vertical clutter and horizontal clutter in linear diagrams, should this be perceived. Formally, the LOS for linear diagrams is the sum of the LS and OS scores for the linear diagram. For example, in Fig. 3, the LOSs are as follows: (v1) 17; (v2) 15; (v3) 14; and (v4) 12.

3 Experiment Design

The study consisted of four tasks, each of which required participants to rank 12 linear diagrams, with a ranking of 1 being least cluttered and 12 being most cluttered. Joint rankings were permitted. The first three tasks fixed the number of sets being represented to 4, 6, and 8 respectively. This allowed us to establish perceived relative clutter when the number of sets did not change. The final task included linear diagrams with 5, 6 and 7 sets (four linear diagrams for each number of sets). This allowed us to establish whether any of the clutter measures were effective at differentiating diagrams with perceived differences in clutter as the number of sets increased. For each task, a primary design feature of the set of 12 diagrams we used is that the different clutter measures give rise to different rankings of the diagrams. This was important for us to gain insight into the relative effectiveness of the clutter measures.

The 12 linear diagrams generated for task 1 were divided into three sets of four diagrams. The four diagrams in each set represented the same information as each other (so they had the same overlaps) but with different layouts. These layouts varied the (horizontal) order of the overlaps and the (vertical) order of the sets. This meant that the line scores and overlap scores varied, whereas the contour score was necessarily fixed. The contour scores did vary between the three sets of diagrams, however. The 12 diagrams were designed to have

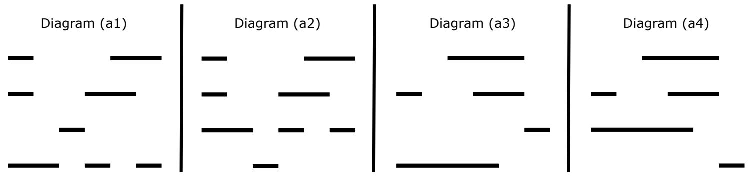

Fig. 4: Linear diagrams with four sets that were used in the study.

three CS clutter levels and six levels of clutter for both of LS and OS (and, consequently, six levels for LOS). Therefore, each clutter measure in the task ranked the 12 diagrams in a different order, allowing us to compare the derived diagram rankings with an overall ranking derived from the participants' rankings. For example, diagrams d1.1 to d1.4 form the first set of diagrams and can be seen in Fig. 3 (note that they were presented to the participants on separate sheets of paper and, as in Fig. 4, without the diagram names being shown). Tasks 2 and 3 have the same design as task one but with 6- and 8-set diagrams.

The diagrams for task 4 consisted of 5-, 6-, and 7-set linear diagrams. Each clutter measure ranked the 12 diagrams in a different order. This was deemed important because the four different rankings allowed us to find out which clutter measure correlates most strongly with the overall participants' ranking. Moreover, for each measure, no pair of diagrams had the same clutter score. This meant that each measure totally ordered the diagrams. For each of the four tasks, table 1 provides details on the diagrams in terms of number of overlaps present and the clutter scores arising from each of the four measures.

4 Experiment Execution, Analysis and Results

Initially eight participants (6 M, 2 F, ages 18-55) took part in a pilot study. The pilot study was successful and the participants finished the four tasks in less than one hour. As no changes were deemed necessary, the pilot data was carried forward for analysis with the data collected in the main study phase. A further 52 participants were recruited, giving a total of 60 participants (46 M, 14 F, ages 18-55). All the participants were staff or students from the University of Brighton; none of them were members of the authors' research group.

To test the effectiveness of the four clutter measures, for each task we first derived an overall ranking of the 12 diagrams from the participants' rankings. Consistent with other researchers who studied diagram complexity, for instance [2, 3, 5], the best clutter measure was identified by the Pearson correlation test on the participants' preference data. We performed a correlation analysis between the overall participants' ranking and that derived from each clutter measure. We viewed a measure of clutter as accurate if there was a significant correlation (at 5%) between the clutter measure ranking and the overall participants' ranking.

For task 1 after collecting all the participants' orderings of the 12 diagrams, we calculated an overall participants' ranking using a Friedman test to estimate

Table 1: The characteristics of the diagrams.

Diagram Number (task 1)	**d1.1**	**d1.2**	**d1.3**	**d1.4**	**d2.1**	**d2.2**	**d2.3**	**d2.4**	**d3.1**	**d3.2**	**d3.3**	**d3.4**
Number of Overlaps	6	6	6	6	8	8	8	8	10	10	10	10
CS Score	11	11	11	11	18	18	18	18	24	24	24	24
LS Score	8	8	5	5	11	11	7	7	11	11	6	6
OS Score	9	7	9	7	12	11	12	11	14	12	14	12
LOS Score	17	15	14	12	23	22	19	18	25	23	20	18
Diagram Number (task 2)	**d1.1**	**d1.2**	**d1.3**	**d1.4**	**d2.1**	**d2.2**	**d2.3**	**d2.4**	**d3.1**	**d3.2**	**d3.3**	**d3.4**
Number of Overlaps	18	18	18	18	22	22	22	22	26	26	26	26
CS Score	60	60	60	60	67	67	67	67	85	85	85	85
LS Score	26	26	14	14	35	35	16	16	42	42	19	19
OS Score	32	25	32	25	41	37	41	37	46	41	46	41
LOS Score	58	51	46	39	76	72	57	53	88	83	65	60
Diagram Number (task 3)	**d1.1**	**d1.2**	**d1.3**	**d1.4**	**d2.1**	**d2.2**	**d2.3**	**d2.4**	**d3.1**	**d3.2**	**d3.3**	**d3.4**
Number of Overlaps	24	24	24	24	30	30	30	30	36	36	36	36
CS Score	103	103	103	103	127	127	127	127	134	134	134	134
LS Score	48	48	24	24	69	69	26	26	68	68	32	32
OS Score	50	46	50	46	70	53	70	53	79	66	79	66
LOS Score	98	94	74	70	139	122	96	79	147	134	111	98
Diagram Number (task 4)	**d1**	**d2**	**d3**	**d4**	**d5**	**d6**	**d7**	**d8**	**d9**	**d10**	**d11**	**d12**
Number of Sets	5	5	5	5	6	6	6	6	7	7	7	7
Number of Overlaps	11	12	13	14	15	16	17	18	19	20	21	22
CS Score	20	28	36	44	40	48	56	64	60	68	76	84
LS Score	15	8	18	10	24	11	29	16	33	19	41	21
OS Score	17	14	23	20	30	26	35	32	44	38	45	48
LOS Score	32	22	41	30	54	37	64	48	77	57	86	69

the median ranking for each diagram. This was then converted the estimates into an overall participants' ranking. Table 2 shows the overall participants' ranking for task 1 in the appropriate row. The results for task 1 are given in the first row of Table 3 which shows the correlation coefficients and, in brackets, the p-values; bold typeface indicates significance. We can see, therefore, that the strongest significant correlation is with the line score. In addition, the line-and-overlap score is significantly correlated whereas the contour score and the overlap score are not.

Table 2 shows the rankings of the 12 diagrams for tasks 2 and 3 alongside the overall participants' ranking. Table 3 shows the correlation coefficients and the p-values. For both tasks, the strongest significant correlation is between the overall participants' ranking and the LS measure, with LOS also being significant.

For task 4, Table 2 shows the rankings of the 12 diagrams for task 4 alongside the overall participants' ranking. Table 3 shows the correlation coefficients and the p-values. As with the other three tasks, the strongest significant correlation is between the overall participants' ranking and the LS measure. However, for this task all clutter measures significantly correlate with the overall participants' ranking.

Table 3 shows that both of LS and LOS measures were significantly correlated to the overall participants' ranking in all four tasks. To establish whether

Table 2: Clutter rankings and participants' ranking for task 1-4.

Task 1	**d1.1**	**d1.2**	**d1.3**	**d1.4**	**d2.1**	**d2.2**	**d2.3**	**d2.4**	**d3.1**	**d3.2**	**d3.3**	**d3.4**
CS Ranking	2.5	2.5	2.5	2.5	6.5	6.5	6.5	6.5	10.5	10.5	10.5	10.5
LS Ranking	7.5	7.5	1.5	1.5	10.5	10.5	5.5	5.5	10.5	10.5	3.5	3.5
OS Ranking	3.5	1.5	3.5	1.5	8.5	5.5	8.5	5.5	11.5	8.5	11.5	8.5
LOS Ranking	4	3	2	1	10.5	9	7	5.5	12	10.5	8	5.5
Participants' Ranking	5	8	2	1	12	10.5	7	6	10.5	9	4	3
Task 2	**d1.1**	**d1.2**	**d1.3**	**d1.4**	**d2.1**	**d2.2**	**d2.3**	**d2.4**	**d3.1**	**d3.2**	**d3.3**	**d3.4**
CS Ranking	2.5	2.5	2.5	2.5	6.5	6.5	6.5	6.5	10.5	10.5	10.5	10.5
LS Ranking	7.5	7.5	1.5	1.5	9.5	9.5	3.5	3.5	11.5	11.5	5.5	5.5
OS Ranking	3.5	1.5	3.5	1.5	7.5	5.5	7.5	5.5	11.5	7.5	11.5	7.5
LOS Ranking	6	3	2	1	10	9	5	4	12	11	8	7
Participants' Ranking	7.5	7.5	1	2	9	10	3	4	11	12	5	6
Task 3	**d1.1**	**d1.2**	**d1.3**	**d1.4**	**d2.1**	**d2.2**	**d2.3**	**d2.4**	**d3.1**	**d3.2**	**d3.3**	**d3.4**
CS Ranking	2.5	2.5	2.5	2.5	6.5	6.5	6.5	6.5	10.5	10.5	10.5	10.5
LS Ranking	7.5	7.5	1.5	1.5	11.5	11.5	3.5	3.5	9.5	9.5	5.5	5.5
OS Ranking	3.5	1.5	3.5	1.5	9.5	5.5	9.5	5.5	11.5	7.5	11.5	7.5
LOS Ranking	6.5	4	2	1	11	9	5	3	12	10	8	6.5
Participants' Ranking	7	8	1	2	9	10.5	4	3	10.5	12	6	5
Task 4	**d1**	**d2**	**d3**	**d4**	**d5**	**d6**	**d7**	**d8**	**d9**	**d10**	**d11**	**d12**
CS Ranking	1	2	3	5	4	6	7	9	8	10	11	12
LS Ranking	4	1	6	2	9	3	10	5	11	7	12	8
OS Ranking	2	1	4	3	6	5	8	7	10	9	11	12
LOS Ranking	3	1	5	2	7	4	9	6	11	8	12	10
Participants' Ranking	4	1	5	2	9	3	10	6	11	7	12	8

Table 3: Correlations between clutter measures and perception, by task.

	CS	LS	OS	LOS
Task 1	0.311 (0.325)	**0.948 (0.000)**	0.374 (0.232)	**0.798 (0.002)**
Task 2	0.474 (0.120)	**0.991 (0.000)**	0.381 (0.222)	**0.851 (0.000)**
Task 3	0.459 (0.133)	**0.942 (0.000)**	0.361 (0.249)	**0.867 (0.000)**
Task 4	**0.615 (0.033)**	**0.993 (0.000)**	**0.839 (0.001)**	**0.958 (0.000)**

LS is significantly more correlated than LOS we used the Fisher r-to-z transformation which converts correlations into a normally distributed measure. Then we used a Z-test to see whether LS is significantly more correlated than LOS. The calculated values of z for the four tasks were as follows: 3.84, 7.69, 2.32, and 4.83 respectively. A one-tailed test yields p-values of 0.0001, 0.0000, 0.0102, and 0.0000 respectively which are all less than 0.05. Therefore the LS measure is significantly more correlated with the overall participants' ranking than the LOS measure.

In summary for each of the four tasks, both the line score and the line-and-overlap score were significantly correlated with participants' perception of clutter in linear diagrams. In each case, however, there was a significantly stronger correlation with the line score. This is unsurprising as, at least for tasks 1 to 3, the overlap score did not yield a diagram ranking that was significantly correlated with the overall participants' ranking. In particular, the overlap score correlation coefficient for these three tasks was quite low, demonstrating that there was little relationship at all. This indicates why adding the overlap score to the line score, yielding the line-and-overlap score, resulted in a weaker correlation.

Recall that task 4 was the only task to include a variety of numbers of sets. This design feature allowed us to gain insight into whether the clutter measures were able to distinguish differences in perceived clutter as the number of sets varied. Interestingly, task 4 (and only task 4) yielded data where all four measures were significantly correlated with perceived clutter. However, the strongest correlation was still with the line score. This indicates that simply comparing the number of line segments present in linear diagrams effectively reflects perceived levels of clutter regardless of the number of sets being visualized: linear diagrams with fewer line segments are perceived to be less cluttered.

5 Conclusion

This paper has provided an understanding of how people perceive clutter in linear diagrams. By considering the syntax of linear diagrams, and how it can be altered through reordering overlaps and sets, we identified four potential measures of clutter, namely: the contour score (generalized from similar research on Euler diagrams), the line score, the overlap score and the line-and-overlap score. Through empirical research, we established that the line score significantly correlates with perceived clutter, regardless of the number of sets present in linear diagrams. In summary, the relative number of line segments present in linear diagrams accurately predicts perceived relative levels of clutter.

The results of our research tell us that reducing the number line segments in linear diagrams reduces perceived clutter. Techniques already exist for reducing the number of line segments, as implemented in the linear diagram generator used to create the diagrams in our study [4]. A key future research goal is to establish the impact of clutter in linear diagrams on user task performance.

References

1. Chapman, P., Stapleton, G., Rodgers, P., Micallef, L., Blake, A.: Visualizing sets: an empirical comparison of diagram types. In: Diagrams, pp. 146–160. No. 8578 in LNAI, Springer (2014)
2. John, C., Fish, A., Howse, J., Taylor, J.: Exploring the notion of 'clutter' in Euler diagrams. In: Diagrams, pp. 267–282. No. 4045 in LNAI, Springer (2006)
3. Purchase, H.C., Freeman, E., Hamer, J.: An exploration of visual complexity. In: Diagrams, pp. 200–213. No. 7352 in LNAI, Springer (2012)
4. Rodgers, P., Stapleton, G., Chapman, P.: Visualizing sets with linear diagrams. ACM Transactions on Computer-Human Interaction (TOCHI) 22(6), 27 (2015)
5. Rosenholtz, R., Li, Y., Nakano, L.: Measuring visual clutter. Journal of Vision 7(2), 17 (2007)
6. Sato, Y., Mineshima, K.: How diagrams can support syllogistic reasoning: An experimental study. JoLLI 24(4), 409–455 (2015)