Interactivity in Linear Diagrams

Peter Chapman[✉][iD]

Edinburgh Napier University, Edinburgh, UK
p.chapman@napier.ac.uk

Abstract. Linear diagrams have been shown to be an effective method for representing set-based data. Moreover, design principles have been empirically developed that, when followed, improve the efficacy of linear diagrams. These principles are task-independent. However, linear diagrams may be produced to aid with a variety of tasks, for which different representations may be more effective. In this paper, we introduce simple interactivity into linear diagrams. Namely, we gave users: the ability to move sets; the ability to move overlaps (set-intersections); and the ability to focus the diagram on a particular group of sets. Whether these interactions improved cognition was investigated via two empirical studies. In the first, we observed that interactivity improved participants' accuracy, confidence and speed. In the second, we observed that these improvements were based on the diagrams participants produced in the first study, rather than being an artefact of interactivity itself. We conclude that adding simple interactivity is useful in the case of linear diagrams.

Keywords: Linear diagrams · Interaction · Empirical evaluation · Set-based representation

1 Introduction

Linear diagrams can be effective at representing set-based data [3]. In contrast to region-based diagrams such as Euler and Venn diagrams, they are also easy to draw. We interpret a linear diagram in the following way: where two horizontal lines exist in the same vertical space, the intersection between the represented sets is non-empty. For example, were a linear diagram to contain some vertical space where lines for set A and B were present, but C was absent, then the diagram would represent that the intersection $A \cap B \cap \bar{C}$ is non-empty.

In essence, a linear diagram can be seen as a matrix, where each matrix entry is either an empty space, or contains a line segment. The columns of the matrix are the *overlaps* of the diagram, and the rows of the matrix are the *sets* of the diagram. With this conceptual model in mind, matrix operations can be recontextualised as linear diagram manipulations, and easily implemented as interactive elements. Making use of these interactive elements could cause the linear diagram to violate drawing guidelines in [15]. Specifically, the number of line segments present in the diagram could increase. The guidelines were

developed for static diagrams, and so have to apply when the task for which the diagram is being used is not known. By contrast, through interaction the user can manipulate the diagram to produce a layout that is helpful for the specific task they have in mind. In this paper, we evaluate whether or not interaction can aid users, when compared to using static diagrams which follow best practice.

In Sect. 2, we investigate interaction, and explain the specific interactive elements added to linear diagrams in Sect. 2.1. The first user-study is explained in Sect. 3, including discussion of the diagrams generated through the interaction process (Sect. 3.5). A question arising from the results of study 1 is investigated in study 2, in Sect. 4. Finally, we conclude and outline further research areas in Sect. 5. The study materials, including diagrams, datasets, and statistical analyses can be found at https://doi.org/10.17869/enu.2021.2748492.

2   Interaction Design

Despite being of central importance to the usability of information visualisations, interaction has until recently received relatively little attention in the literature [7]. Where new visualisations are introduced, the interactive aspects are treated as of secondary importance, if at all [18]. Interaction is being increasingly studied, however, in the literature [13]. Interactions themselves need not be complex, as Dix and Ellis state: "virtually any static representation can become more powerful by the addition of simple interactive elements" [6]. We seek to test this (theoretical) assertion in an empirical manner in this paper.

The literature on interaction in the InfoViz field which does exist is concerned with classifying broad types of interaction users might want. These types are categorised differently by different authors. For example, Shneiderman [16] provides seven tasks that users may wish to undertake: *overview, zoom, filter, details-on-demand, relate, history* and *extract.* Dix and Ellis [6], meanwhile, suggest that the necessary interactions are: *highlighting and focus; accessing extra information; overview and context; same representation, changing parameters; same data, changing representation;* and *linking representations.*

The approach of [7] is different. Rather than focussing on what tasks interaction should support, Elmqvist *et al.* provide a set of eight interaction guidelines for fluidity. When followed, the authors contend that they will produce "effective information visualizations that support fluid interaction." In the following section, we explain the interaction elements added to linear diagrams, and relate them to the theoretical guidelines from the literature.

2.1   Interactive Elements in Linear Diagrams

We introduce four interactive controls to a linear diagram, extended from [4]. Two concern the horizontal ordering of the overlaps, and two concern the vertical ordering of the sets. All alter the diagrams in a sound manner: the underlying information represented is not changed through these interactions [5].

Swap Two Adjacent Overlaps. An example of a button which controls this interaction can be seen in the circle labelled 1 in Fig. 1.

Force Order. This element is the most complex interaction. For this paper, we use one that can prioritise up to 2 sets. The 2-set prioritisation works as follows. The user supplies up to two sets, for example A and B. The overlaps are then split into four mutually exclusive groups, namely those containing: A but not B; both A and B; B but not A; and finally neither A nor B. If only one set is selected, then the overlaps are split into two groups: those containing the set selected, and those not containing the set selected. The diagram is then drawn by separately ordering the four groups of overlaps using the drawing algorithm of [15] and then concatenating them together. In this way we can guarantee that the lines A and B (in the two-set case) will be drawn using single line segments. How this interactive element appears is seen at 2 in Fig. 1. An example of the outcome can be seen in Fig. 2, where the order was forced on the circled sets.

Swap Two Adjacent Sets. As an example, if the button in the circle labelled 3 in Fig. 1 was clicked, then the sets Economics and Food would be switched. The sets would keep their original colours: i.e. Economics would still be purple after the switch, whilst Food would still be brown. By maintaining the original colours, a user's mental map can be preserved [12].

Move Set to Top. This interaction is encoded as repeated application of the "swap two adjacent sets" functionality. The user does not see the intermediate steps, however: the new set simply appears at the top. The remainder of the sets keep their colours and relative order. As an example, if the button in the circle labelled 4 in Fig. 1 was clicked, then the set Movie would move to the top, and all other sets would move down by one.
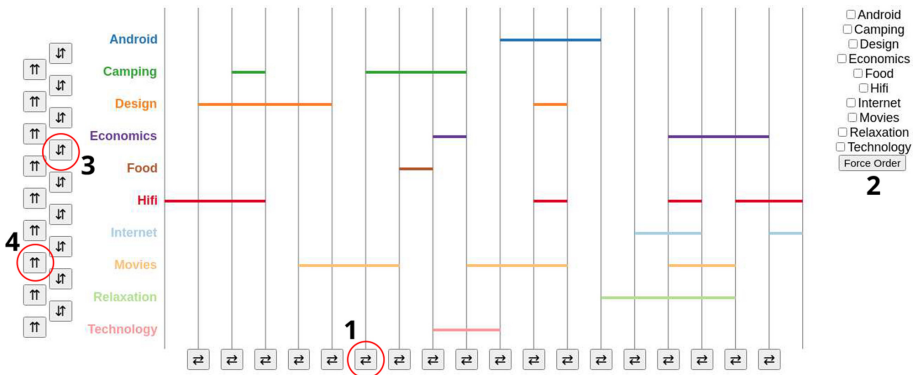

Fig. 1. Interactive elements for linear diagrams: 1. swap overlaps; 2. force order; 3. swap sets; 4. move set to top
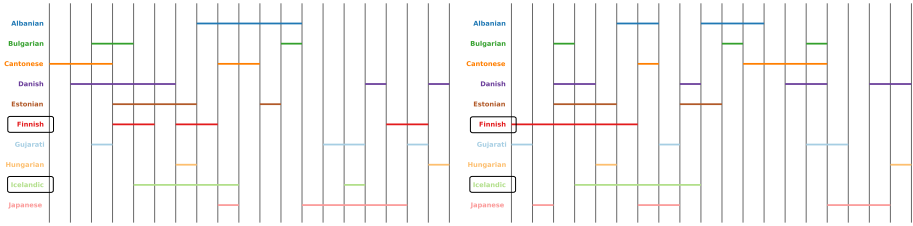
Fig. 2. Original (left) vs. forced (right): forcing on Finnish and Icelandic.

Discussion. The interactions outlined above change the representation of the data, but not the data itself. From the framework of Shneiderman [16], the tasks that will be possible with these interactions are *relate* (view relationships among items) and *details-on-demand* (select an item or group and get details when needed). To some extent, the interactive elements also support *zoom* (zoom in on items of interest) and *filter* (filter out uninteresting items); however the altered diagrams do not remove the uninteresting information (or zoom onto the interesting information) but can rather partition the diagram into interesting and uninteresting regions.

From the framework of Dix and Ellis [6], meanwhile, the kinds of interaction supported are *highlighting and focus* along with *same representation, changing parameters*. In the former, by bringing the overlaps/sets of interest into close proximity to each other, we exploit that "interaction is good, but eye movements [...] are faster still" [6]. For the latter, forcing the drawing order of the diagram can possibly allow more effective representation of parts of the dataset.

Of the eight design guidelines outlined by Elmqvist [7], some were not attempted. As this work is the first step towards an empirically validated tool, it was determined that the use of animated transitions between diagrams was not of high importance. Thus design guidelines 1 (use smooth animated transitions between states) and 5 (reward interaction, through "the use of animations, sounds and pretty graphics") were not followed. (This aspect will be revisited in Sect. 5.) However, the interactions endeavoured to follow design guidelines 2 (provide immediate visual feedback on interaction); 3 (minimize indirection in the interface) and 4 (integrate user interface components in the visual representation) by having the buttons for manipulating overlaps be directly adjacent to the diagram, and not separated by any border etc.; and 7 (reinforce a clear conceptual model) by maintaining the first-presented colours of each set, regardless of where the set ends up after manipulation. Of secondary importance were the guidelines 6 (ensure that interaction never 'ends') and 8 (avoid explicit mode changes). For the former, there is always the ability to continue interacting with the diagram, whereas for the latter there was only one representation type present, and so changing mode (i.e. representation type) was impossible.

Our interactive elements are relatively narrow in their scope, but the representation itself is also limited to displaying set-based information. As such, the interactive elements should help the user complete most of the tasks associ-

ated with set-based data (i.e. determining the intersections, unions, disjointness, containment etc. between various groups of sets). The interactions follow suggested guidelines, with the biggest omission that of animated transitions. In the next section, we determine whether the tools which were designed to help users complete a range of tasks are actually helpful with one specific task.

3 First Study: Interactive vs. Static Diagrams

The research question we will be answering is:

> [**RQ1**]: Does interaction aid users in performing set-based tasks with linear diagrams?

Owing to space limitations, we restrict the number of independent variables considered. We always use diagrams that contain 10 sets, and either 19 or 20 overlaps. We also only focus on one particular task, with two variants: identifying intersections between a set and a given combination of two other sets. The variants come from considering two different given combinations of sets. These choices necessarily limit the direct applicability of the results to other tasks, and other size diagrams. However, we can still address our research question.

Examples of the two task variants are to identify the sets where "**some** of the set is also in common with **either** set A **or** set B", and to identify the sets where "**some** of the set is also in common with **both** set A **and** set B". In other words, in the first we are asking participants to identify those sets which intersect with $A \cup B$, and in the second we are asking participants to identify those sets that intersect with $A \cap B$. Given that the use of multiple line segments to represent a set is permissible in linear diagrams, both $A \cup B$ and $A \cap B$ may be relatively concentrated in one small area of the diagram, or spread widely across it. In order to control for this variance, we investigate the concept of *question shape*.

Question Shape. Participants will be required to interrogate a diagram to determine whether or not a given combination of sets has elements in common with remaining sets. In order to do this participants *must* find the given combination of sets described in the task. Once this region of the diagram is discovered, the participant should focus their attention on this region, and expand their focus if needed. By controlling the size of this task region (hereafter referred to as *question shape* or just *shape*), we can determine whether various measures of diagram performance are affected by interactivity and shape.

Shape is determined in two dimensions. Within the context of linear diagrams, a tall shape would represent a situation where the question sets are separated by some vertical distance. By contrast, a short shape would represent a situation where the question sets are relatively close vertically. A wide shape would represent a situation where the question overlaps span a large horizontal distance. Finally, a narrow shape would represent a situation where the question overlaps are relatively close horizontally. Taken together, these two dimensions

give rise to four shapes: a short-wide shape (hereafter represented in the paper as ⊑); a short-narrow shape (□); a tall-narrow shape (▯); and a tall-wide shape (▢). The four question shapes can be seen highlighted in Fig. 3.

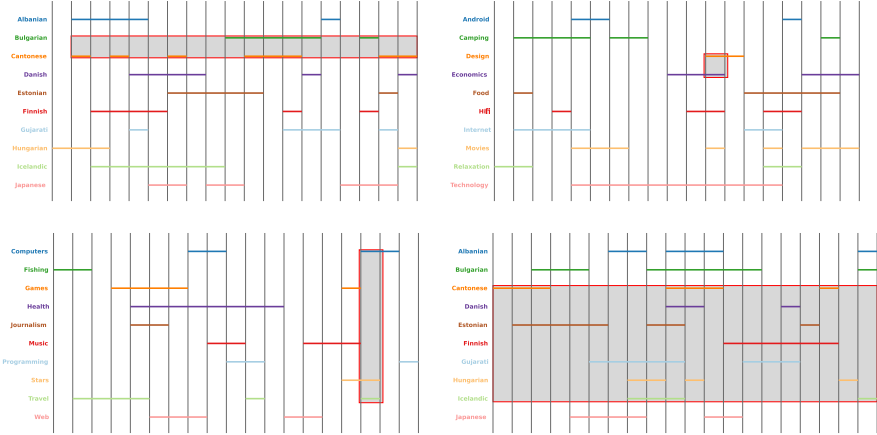**Fig. 3.** Question shapes: Top-left the question refers to Bulgarian ∪ Cantonese; top-right Design ∩ Economics; bottom-left Fishing ∩ Travel; bottom-right Cantonese ∩ Icelandic.

We characterise a narrow shape as one which spans less than half of the horizontal space of the diagram, and a wide shape as one which spans greater than half of the horizontal space of the diagram. Similarly, a short shape is one which spans less than half of the vertical space of the diagram, and a tall shape is one which spans greater than half of the vertical space of the diagram. Three tasks were presented to participants for each shape, yielding 12 tasks in total.

3.1   Study Design and Materials

Our first study has two independent variables: **group** (interactive group vs. static group), and **shape**. We use a between-groups design, in that participants either saw only diagrams they could interact with (interactive group), or only static diagrams (static group). Participants are presented with 12 tasks, with 3 tasks for each shape. An example of how a task was presented is given in Fig. 4, which is task 5 from the study. This figure shows a union variant of a task, with wording "Tick the checkboxes where **some** of the people are also interested in **either** Food **or** Internet". The wording for the task for the other variant would be (for example) "Tick the checkboxes where **some** of the people are also interested in **both** Android **and** Design." Two contexts were used: the interests of a group of people (as seen in Fig. 4), and the languages spoken by a group of people (as seen in the top-left panel of Fig. 3). Because all diagrams in the study have 10 sets, and all tasks follow the pattern seen in Fig. 4 with two given sets, there are

always 8 checkboxes available for selection. The response variables collected for both groups were the number of check-boxes correctly filled in (giving a score from 0 to 8); the self-reported confidence level of the participant (from 0 to 5); and the time taken to answer the question (in milliseconds). In addition, the diagram created by the interactive group was also collected.
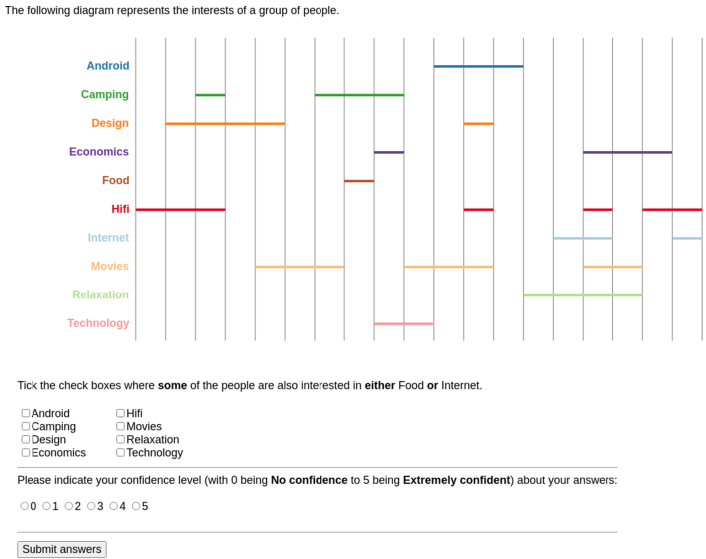

Fig. 4. How a task in the static group appears.

3.2   Hypotheses

The interactive elements will allow participants to redraw a diagram so that the area they must focus their attention on is (a) contiguous, in that the question sets are drawn as single lines, and (b) in a known location on the diagram (at the left-hand side). These two facts should allow participants to find the intersecting sets more easily, and to be more certain that they have found all intersecting sets. However, interacting with the diagram takes time: users must perform two sub-tasks. They must first re-arrange the diagram, and then interrogate the new diagram. We thus have our first hypothesis (in three parts):

[**H1.i/ii/iii**] the use of interaction leads participants to answer *significantly more accurately/confidently/slower* than when using static diagrams.

When considering shape, the more space a user needs to interrogate, the more potential there is for mistakes. Further, it could take longer. There is a possible

interaction between shape and interaction: as noted above the use of interaction can reduce the search space for the user. However, we could still reasonably expect larger question shapes to be more challenging to users. We thus have our second hypothesis (in three parts):

[**H2.i/ii/iii**] diagrams with a small question shape permit users to answer tasks *significantly more accurately/confidently/quickly* than with diagrams with a large question shape.

3.3   Methods

Data Collection. Participants were recruited from a university in the southeast of Scotland. Owing to the pandemic, face-to-face data collection was not possible. Participants were invited to download the study materials (when packaged as a zip), and open the first page (`start.html`) in a browser. The study would then run locally on their machine. A random number was generated on the start page: this was used to assign participants to a group, and to generate a unique identifier. Only the participant knows both parts of the identifier-identity pair. The responses were automatically collected in a text file, which was produced at the end of the study. Participants were then asked to submit the text file to an `ftp` server. Once collected, the files were deleted from the server. This study received ethical approval from Edinburgh Napier University.

Training. There were two phases to the study: a training phase and a main phase. In the training phase, participants were first given a description of a linear diagram itself, and how to interpret them. For the interactive group, they were further given a description of what each interactive control did. Note that they were not given any information of *when* to use a control, but rather only *how* to use a control. In other words, the interactive controls were not linked to the specifics of completing a task.

Participants were then presented with three training questions to familiarise them with the task. In the first and second, the correct answers were pre-selected, and explained. In the third, the correct answers were not pre-selected. Participants were then taken to a holding page, which contained information about how many questions were remaining, and given the opportunity for a rest.

Main Phase. Participants were presented the 12 tasks in a random order. Participants were free to select no sets as their answer. However, participants were not able to submit their answer without selecting a confidence-level. Between consecutive tasks, users were taken to a holding page where they could have a rest. The number of questions remaining was presented on this page. The holding page also fulfilled a technical role: various timers and variables were reset on the holding page. After all tasks were completed, participants were asked to indicate their age (in bands) and whether or not they had any colour-blindness. Responses of "prefer not to say" were available for both questions.

Statistical Analysis. Multiple responses were collected from each participant. The dataset therefore exhibits clustering, meaning that the assumption of standard approaches (ANOVA, χ^2-test for goodness of fit, etc.) of independence of observations was violated, possibly leading to overstated statistical significance and underestimated standard errors [2]. We thus used the approach of generalised estimating equations [8], used in conjunction with generalised linear models (GLM). The statistical software `R` was used for the analysis, with the package `geepack` [9]. This approach still yields *p*-values, and so is appropriate for hypothesis testing. The models fitted are explained in each section. Initially, a model with both main terms (here shape and group) and interaction terms was attempted, and simplified if no interaction effect was present.

3.4   Interaction vs. Static Results

A pilot study was conducted, with 3 participants in each group (interactive, and static). No issues were identified during the pilot, and no changes to the study design were made, and so these 6 participants were included in the main results. A total of 53 participants were recruited, of which 29 were randomly assigned to the static group, and 25 to the interactive group. As the participants were students on Computing-related courses, the sample skewed male and young.

It was observed that, of the 25 participants in the interactive group, not all of them used the interactive tools available to them. 17 of the 25 made changes to the diagrams presented (the details of these changes will be discussed in Sect. 3.5). Of note is that participants either altered all 12 diagrams they saw, or altered none of them. From an analysis point of view, a decision thus needed to be made. Either the groups remained an independent variable in the model (i.e. the distinction between participants is whether they had the *potential* to interact with the diagrams, regardless of whether they *did*), giving a group split of 25 to 29; a new independent variable is introduced to encode whether a participant interacted with the diagrams or not, giving a group split of 17 to 37; or the non-interacting participants in the interactive group were discarded, giving a group split of 17 to 29. Because we are interested in whether or not the interactions participants make are useful, rather than whether the interface encouraged interaction, we have split the data according to the middle choice. Thus, a new independent variable (changed vs. unchanged) was introduced.[1]

Accuracy. Both groups had high accuracy rates. Each question contributed a score of 0 to 8: the average question score in the changed group was 7.83, and in the unchanged group was 7.61 (giving accuracy rates of 98% and 95% respectively). Whilst high, these rates are not unexpected. High rates were found in [3,15]. For each shape, the following accuracy rates were observed: for the shape ▭, the rates were 94.8% changed vs. 95.7% unchanged; for □, 99.5% changed vs.

---

[1] The results remain true whichever choice was made. The other models can be seen in https://doi.org/10.17869/enu.2021.2748492. The *p*-values differ, but the models do not give remove (or introduce) significance at 95%.

96.3% unchanged; for ⬜, 99.5% changed vs. 94.7% unchanged; and for ⬜, 97.9% changed vs. 93.6% unchanged.

A GLM with an ordinal response variable was initially anticipated. However, owing to the high accuracy rates some categories did not have enough responses for robust analysis [10].[2] Thus, responses were combined in the following way: question responses were recoded as either the participant was completely correct (i.e. a score of 8 on a question), or made at least one mistake (i.e. a score of 7 or lower). In this way, the model became a GLM with a binomial response. It is these results which we report. The model gives an interaction effect between the group and shapes ⬜ ($p = 0.0195$) and ⬜ ($p = 0.0153$). We can thus conclude that for shapes ⬜ and ⬜, *participants who changed the diagrams performed significantly better than those who did not.* For the other two shapes ⬜ and ⬜, there was no difference between the two groups.

Confidence. Both groups had high confidence rates. The changed group reported an average confidence level of 4.7 (out of 5), and the unchanged group reported an average confidence level of 4.4. By shape, we observed the following confidence levels. For ⬜, 4.7 for changed vs. 4.2 for unchanged; for ⬜, 4.9 for changed vs. 4.5 for unchanged; for ⬜, 4.8 for changed vs. 4.5 for unchanged; and for ⬜, 4.6 for changed vs. 4.2 for unchanged.

As with the accuracy results, there were not enough responses in some categories to perform robust ordinal regression, and so a GLM with a binomial response was fitted. The model thus compares those who had full confidence (i.e. level 5) with those who had confidence levels of 4 or lower.

There was no interaction effect between group and shape, and so a more simple model was fitted with only main terms for group and shape. The changed group were found to be *more confident* in their responses than the unchanged group ($p = 0.035$). Further, for the narrow shapes ⬜ ($p = 0.026$) and ⬜ ($p < 0.001$), participants were more confident with their responses than with the wide shapes ⬜ and ⬜. No differences were found between ⬜ and ⬜, or between ⬜ and ⬜.

Time. The average time taken for the changed group to answer a question was 45.0 s, compared with 63.3 s for the unchanged group. Note that this time *includes* any time re-arranging the diagram for the changed group. When comparing by shape, we have: for ⬜, 52.0 s for changed vs. 80.6 s for unchanged; for ⬜, 39.3 s for changed vs. 47.6 s for unchanged; for ⬜, 33.4 s for changed vs. 47.2 s for unchanged; and for ⬜, 55.3 s for changed vs. 77.8 s for unchanged.

A GLM with a normal response was fitted to the data. There was found to be no interaction effect between shape and group, and so a simpler model with just main terms for group and shape was fitted. It was found that the changed group were *significantly faster* ($p = 0.0073$) than the unchanged group, and that shapes ⬜ ($p < 0.0001$) and ⬜ ($p < 0.0001$) were *significantly faster* than the

---

[2] These models were developed, and broadly show the same results as the binomial model. However, the more conservative model is reported as it is more robust. The ordinal logistic model can be found with the rest of the study materials at https://doi.org/10.17869/enu.2021.2748492.

shapes ⬭ and ☐. There were no differences within each pair of shapes: neither the narrow shapes (□ and ▯) nor the wide shapes (⬭ and ☐).

Discussion. We can answer RQ1: interaction can help users in performing set-based tasks with linear diagrams. We have seen that hypotheses **H1.i** and **H1.ii** can be supported by the data. However, **H1.iii** can be rejected. The evidence is more mixed for **H2**. The tall shapes produced worse accuracy results (evidence for **H2.i**), whereas the narrow shapes gave higher confidence **H2.ii** and quicker responses **H2.iii**. Overall, larger shapes were more problematic, but the dimension of size increase was important.

The high levels of accuracy amongst participants, as mentioned, was not unexpected. We also gained further (anecdotal) evidence that participants struggle with broken lines, as the wide shape questions produced lower confidence. This phenomenon has been observed before: in [15], which recommended minimising the number of line segments in a diagram; in [1], more line segments equated to a higher perceived clutter; and in [17] participants reported that "broken lines were problematic". Of interest here, however, is that whilst participants were less confident with finding information across a wide area of the diagram, there was no significant lowering of accuracy in these cases. Indeed, what was found to increase the accuracy was to reduce the *vertical* size of the question, not the horizontal. In some cases, it would not be possible to change the horizontal shape (from wide to narrow), but it was always possible to change the vertical shape (from tall to short). We will return to this theme in Sect. 3.5.

We see that confidence levels for those who made changes are higher even when there is no corresponding improvement in accuracy (i.e. for shapes ⬭ and □). This finding could be seen as an example of the illusion of control [11], where increased confidence in performance is not related to an improved performance when participants have perceived control. In general, however, it is re-assuring that the overall high levels of accuracy translate into high levels of confidence.

The time findings were surprising. Participants were *faster* when they had to perform two tasks rather than one: first alter the diagram, and then interrogate the new diagram to produce the required answer. Of course, participants who changed the diagram then knew where the information they were looking for was to be found. Thus, there would be little need to search the diagram, or to spend time checking other regions of the diagram that were known to be outside the area of interest.[3] However, that no searching outside a given region would be necessary is known only to the user who altered the diagram. It is an open question as to whether the benefits of the layout are available to users who did not generate the layout. We attempt to answer this question in the second study, in Sect. 4.

---

[3] This was alluded to as the zoom-like or filter-like behaviour of [16] in the Discussion of Sect. 2.1.

3.5   User-Generated Diagrams

There were 17 participants who changed diagrams, as detailed in Sect. 3.4. In this section, we examine those altered diagrams. Owing to space restrictions, we focus on the diagrams created for shapes ⧠ and ⊑; these are the shapes where participants saw an accuracy improvement. However, the themes hold for the other shapes, too.

Two main patterns for redrawing were observed. The first can be seen in the left panel of Fig. 5, which was a diagram created by a participant to help complete task 4 (originally shown in the lower-right panel of Fig. 3): *tick the check boxes where* **some** *of the people also speak* **either** *Cantonese* **or** *Icelandic.* The participant has used the "force order" functionality, applied to the lines Cantonese and Icelandic, but has left those two lines in their original vertical location. As such, they have transformed what was a ⊑ diagram (the lower-right figure of Fig. 3) into a ⧠ diagram (Fig. 5, left).
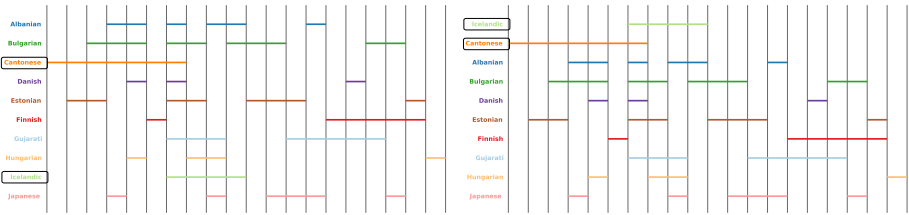

Fig. 5. User-generated diagrams: forced order only (left), forced order followed by vertical re-ordering (right). The sets of interest are circled.

The second main pattern observed can be seen in the right panel of Fig. 5. Again, this diagram was collected from a participant's interaction for the same task. The participant has used the "force order" functionality, but has then subsequently moved the two question lines to the top of the diagram. As such, they have transformed what was a ⊑ diagram into a ⊓ diagram.

For the six tasks concerning ⧠ and ⊑ diagrams, the frequency of each pattern is given in Table 1. The category "Other" includes: instances where force-order was applied to only one set; where force-order was applied to both and some vertical re-arrangement made the question sets be adjacent but not at the top; and one instance some individual overlaps had been moved. Within the "force order, move to top" category, there were two subcategories. The first is seen in the right panel of Fig. 5: the line which starts left-most is *below* the other. The second is not shown for space reasons, but transposes the top two sets. Participants produced the former over the latter at a ratio of roughly 3:1.

Discussion. We can infer from the frequencies given in Table 1 exactly which functionality participants used. The force order functionality was used in all but one of the 102 diagrams created. The ability to move a set to the top of the

Table 1. Frequency of user-generated diagrams

Task	Force order	Force order, move to top	Other
3	3	13	1
4	4	13	0
7	3	13	1
8	5	12	0
11	4	13	0
12	3	12	2

diagram was used in 75% of instances. However, the ability to move a single overlap left and right, and the ability to move an individual set up and down one step at a time, were rarely, if ever used. It would be beneficial to remove that functionality: the buttons are creating visual clutter on the screen, but not performing any useful function for this task.

4   Second Study: User-Generated vs. Original Diagrams

The results of Sect. 3.5 lead us to pose the following research question:

> [**RQ2**]: are the improvements in the interactive group owing to the user-generated layout, or the process of interaction itself?

In other words, if person A created a diagram through interaction, would person B gain the performance benefits if shown that diagram over the original? We could address this question partially through the analysis of Sect. 3.4. If the process of interaction was driving the accuracy improvements, then we would have expected to see uniform improvements for all shapes. This uniform improvement did not occur, although there was a confidence improvement.

We can also address the question through a secondary study, however. By presenting each participant with both the original and the most common user-generated diagram (i.e. Fig. 5, right), we can observe whether accuracy, confidence and time follow the same changes as for those who created the diagrams. We only present participants with the diagrams that were found to improve accuracy in Sect. 3.4, namely ⎕ and ⬚ shaped diagrams, in order to keep the anticipated time per participant to complete the set of tasks low enough so as not to cause fatigue [14]. Further, we changed the context of the altered diagrams so that participants would not know they were seeing each task twice: i.e. if the original diagram was about languages, then the user-generated diagram was altered to be about interests. The alphabetical order was retained, however.

We thus conducted a second study with a *within* factor (the **type** of the diagram, with two levels user-generated and original). In addition, we had the **shape** factor, with two levels ⎕ and ⬚ within the model. As in Sect. 3.3, there was also an interaction term in the model type × shape. Where there was found

to be no interaction effect, the model was simplified to only include the main terms. The number of correct responses, the confidence level, and the time taken to answer a question was recorded for each participant.

4.1 Hypotheses

The user-generated diagrams contain more line segments than the originals (on average 11.7% more line segments). However, the region of interest for the task is contiguous. A further confounding issue is that the order of the sets is no longer alphabetical: it may be more difficult to find the particular question sets than for the original diagrams. We still anticipate that the contiguous task region will be the most important factor, leading to our hypotheses **H3** (shown later).

With regards to shape, we are only looking at those shapes which caused more problems from the first study. However, we still have two different sizes in the horizontal direction: narrow and wide shapes. Given the results of the first study, we can hypothesise that a similar effect will be present, leading to hypotheses **H4**.

> **[H3.i/ii/iii]** user-generated diagrams permit participants to answer *more accurately/confidently/quickly* than the original diagrams.
> **[H4.i/ii/iii]** narrow shapes permit participants to answer *more accurately/confidently/quickly* than the wide shapes.

4.2 Results

A pilot study was conducted, with 3 participants. No issues were identified during the pilot, and no changes to the protocols were made, and so the pilot participants' data were included in the main dataset. A further 17 participants were recruited. As with the first study, the participants were students enrolled on Computing-related courses, and so skewed male and young.

Accuracy. Participants again showed high accuracy rates. Overall, the user-generated diagrams returned an accuracy rate of 92.4%, whereas the original diagrams had an accuracy rate of 89.3%. When considered by shape: for ⬭, the accuracy was 97.3% for user-generated vs. 94.0% for the original diagrams. For ▢, the rates were 87.5% for user-generated vs. 84.6% for the original diagrams.

As with study 1, the high accuracy rates in some categories meant that responses were recoded as either completely correct, or at least one mistake. The results reported are for a GLM with a binomial response variable. There was no interaction effect between shape and type ($p = 0.778$) and so a simpler model with no interaction terms was fitted. There was a main effect for user-generated vs. original diagrams ($p = 0.0033$), but there was no effect owing to shape ($p = 0.20$). We can thus conclude that *the user-generated diagrams performed significantly better than the original diagrams.*

Confidence. Participants had high confidence scores. The average confidence was 4.5 for the user-generated diagrams, versus 4.2 for the original diagrams. By

shape, the confidence for ⬚ was 4.8 for user-generated vs. 4.4 for original. For ☐, the scores were 4.1 for user-generated vs. 4.0 for the original diagrams.

As with study 1, it was necessary to recode confidence as either full confidence (level 5), or a confidence level of 4 or lower. A GLM with a binomial response was fitted to these data. An interaction effect between type (user-generated vs. original) and shape was present ($p = 0.0041$), and so each type was analysed separately. For the user-generated diagrams, there was significantly lower confidence with shape ☐ than with shape ⬚ ($p = 0.0003$), whilst for the original diagrams there was no difference in confidence levels between the shapes ($p = 0.18$). Overall, there was a significant decrease in confidence between the user-generated and the original diagrams ($p = 0.0004$). Again, we can conclude that *the user-generated diagrams gave participants higher levels of confidence than the original diagrams.*

Time. One participant took roughly 100 times longer on one task than the next highest time. That response was then removed as an obvious outlier. The summary statistics and analysis are described for the reduced dataset. Overall, the average time taken to complete a task was 50.6 s for the user-generated versus 78.8 s for the original diagrams. By shape, we have for ⬚: 33.6 s for the user-generated vs. 60.4 s for the original diagrams. For c, we have 67.4 s for the user-generated vs. 97.2 s for the original diagrams.

A GLM with a normal response variable was fitted to the data. There was no interaction between type and shape ($p = 0.80$), and so a simpler model with no interaction terms was fitted. Participants answered significantly faster ($p < 0.0001$) when answering tasks with user-generated diagrams than with the original diagrams. In addition, ☐ shaped-questions took significantly longer to answer than ⬚ shaped-questions ($p = 0.0085$). We can again conclude that *the user-generated diagrams outperformed the original diagrams.*

Discussion. Participants were more accurate, faster, and more confident, when using user-generated diagrams than the original diagrams, giving supporting evidence for hypothesis **H3**. We can then answer research question **RQ2**: the user-generated layouts give performance enhancements to those who did not create them. Of course, the only diagrams examined are those which produced an improvement, namely ⬚ and ☐. As with the first study, the hypotheses regarding shape are more subtle. For accuracy, we can reject **H4.i**; for confidence, the interaction effect means that we only have partial evidence for **H4.ii**; and for time we have supporting evidence for **H4.iii**.

5   Conclusions and Further Work

We have implemented simple interactive controls for linear diagrams, and shown that they are useful for one type of set-based task. Further, we have identified that the question shape affects user accuracy, confidence and speed. However, the type of shape change is important: confidence varies with horizontal changes,

whereas accuracy and speed vary with vertical. In either instance, interactions which allow the shape of the question to be altered are beneficial.

There is much scope for further work. On an immediate level, the interactive elements themselves could be incorporate smooth transition effects, allowing more guidelines from [7] to be implemented. In more general terms of interaction and linear diagrams, more task types could be investigated. For example, we only looked at intersection. Whether interactivity also helps with containment or disjointness would be an obvious extension. Similarly, moving to a different representation could yield different results. The force order algorithm works by letting the user, rather than a heuristic, specify the drawing order. Where set representations allow a set-by-set drawing algorithm, order forcing through interaction can be implemented.

Region-based representations can draw more sets before known problems occur: whilst linear diagrams can only draw Venn-2 using a single line for each set, it is possible to draw Venn-3 using circles, and Venn-5 using ellipses, without using duplicate curves. A force order algorithm applied to these representations can guarantee reasonable representations for a higher number of sets, and interaction could therefore be highly useful to them.

References

1. Alqadah, M., Stapleton, G., Howse, J., Chapman, P.: The perception of clutter in linear diagrams. In: Jamnik, M., Uesaka, Y., Elzer Schwartz, S. (eds.) Diagrams 2016. LNCS (LNAI), vol. 9781, pp. 250–257. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42333-3_20
2. Cameron, A.C., Miller, D.L.: Robust inference with clustered data. Technical Report, Working paper 10–7, University of California, Davis (2010)
3. Chapman, P., Stapleton, G., Rodgers, P., Micallef, L., Blake, A.: Visualizing sets: an empirical comparison of diagram types. In: Dwyer, T., Purchase, H., Delaney, A. (eds.) Diagrams 2014. LNCS (LNAI), vol. 8578, pp. 146–160. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44043-8_18
4. Chapman, P., Roberts, W.: Towards diagram-based editing of ontologies. In: Chapman, P., Stapleton, G., Moktefi, A., Perez-Kriz, S., Bellucci, F. (eds.) Diagrams 2018. LNCS (LNAI), vol. 10871, pp. 699–703. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91376-6_62
5. Chapman, P., Stapleton, G., Rodgers, P.: PaL diagrams: a linear diagram-based visual language. J. Vis. Lang. Comput. **25**(6), 945–954 (2014)
6. Dix, A., Ellis, G.: Starting simple: adding value to static visualisation through simple interaction. In: AVI 1998, pp. 124–134. ACM (1998)
7. Elmqvist, N., Moere, A.V., Jetter, H.C., Cernea, D., Reiterer, H., Jankun-Kelly, T.: Fluid interaction for information visualization. Inf. Vis. **10**(4), 327–340 (2011)
8. Galbraith, S., Daniel, J.A., Vissel, B.: A study of clustered data and approaches to its analysis. J. Neurosci. **30**(32), 10601–10608 (2010)
9. Halekoh, U., Højsgaard, S., Yan, J., et al.: The R package GEEPACK for generalized estimating equations. J. Stat. Softw. **15**(2), 1–11 (2006)
10. Harrell, F.E., Jr., Lee, K.L., Mark, D.B.: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med. **15**(4), 361–387 (1996)

11. Langer, E.J.: The illusion of control. J. Pers. Soc. Psychol. **32**(2), 311 (1975)
12. Misue, K., Eades, P., Lai, W., Sugiyama, K.: Layout adjustment and the mental map. J. Vis. Lang. Comput. **6**(2), 183–210 (1995)
13. Munzner, T.: Visualization Analysis and Design. CRC Press, Boca Raton (2014)
14. Purchase, H.: Experimental Human Computer Interaction: A Practical Guide with Visual Examples. Cambridge University Press, Cambridge (2012)
15. Rodgers, P., Stapleton, G., Chapman, P.: Visualizing sets with linear diagrams. ACM Trans. Comput. Hum. Interact. (TOCHI) **22**(6), 27 (2015)
16. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: IEEE Symposium Visual Languages, pp. 336–343 (1996)
17. Stapleton, G., Rodgers, P., Touloumis, A., Blake, A.: Well-matchedness in Euler and linear diagrams. In: Pietarinen, A.-V., Chapman, P., Bosveld-de Smet, L., Giardino, V., Corter, J., Linker, S. (eds.) Diagrams 2020. LNCS (LNAI), vol. 12169, pp. 247–263. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-54249-8_20
18. Yi, J.S., Ah Kang, Y., Stasko, J., Jacko, J.A.: Toward a deeper understanding of the role of interaction in information visualization. IEEE TVCG **13**(6), 1224–1231 (2007)