A healthy critical attitude: Revisiting the results of a graph drawing study

Helen C. Purchase [1]

[1]The University of Glasgow

Abstract

This paper reports on a series of three similar graph drawing empirical studies, and describes the results of investigating subtle variations on the experimental method. Its purpose is two-fold: to report the results of the experiments, as well as to illustrate how easy it is to inadvertently make conclusions that may not stand up to scrutiny. While the results of the initial experiment were validated, instances of speculative conclusions and inherent bias were identified. This research highlights the importance of stating the limitations of any experiment, being clear about conclusions that are speculative, and not assuming that (even minor) experimental decisions will not affect the results.

1   Introduction

User experiments are a relatively recent phenomenon in the area of graph drawing research, which has typically focused on the computational and algorithmic aspects of depicting relational information visually. The Graph Drawing Symposium in 1995 included two such studies, the first examples of experimental studies in this area [2, 20].

Research studies on the effectiveness of graph layout algorithms from a user comprehension perspective have increased in recent years, covering, for example, dynamic graphs [1], eye-tracking studies [15], and application areas such as sociograms [12] and software engineering diagrams [26].

Building on empirical methods used in experimental psychology, human-computer interaction, and information visualization, graph drawing experimental studies are typically based on the standard scientific method of changing independent variables, controlling potentially confounding factors, subjecting the stimuli to a well-defined test, and measuring dependent variables.

The model for such experimental research tends to be that researchers run an experiment with human participants, analyse the data, form conclusions, and then publish them: the experimental results are then available to be cited in other researchers' papers, releasing these other researchers from the need to investigate the questions themselves.

Seldom are the results of a published graph drawing experiment subject to further detailed investigation – either by the researcher who first conducted the experiment, or by others. This is despite the fundamental principle of the scientific method that "we do not . . . accept [our observations] as scientific observations until we have repeated and tested them" [16, p23], and that it is common practise in the life sciences [24] and physical sciences [3] for experiments to be replicated to validate results. It is also well known that all human experiments have limitations and constraints, meaning that no experiment is ever perfect [10].

For a typical example from the broader area of information visualization, consider the experiment to investigate whether understanding changes in statistical data presentation is affected by the means by which the transition is depicted [9]. The authors concluded that "animated transitions can significantly improve graphical perception." This experiment has not been replicated: none of the papers citing this paper conducted a similar experiment to validate these results. Despite this, the paper has been cited several times with its experimental conclusion of this paper stated as fact: "Transitions from one query to the next are smoothly animated to preserve the user's sense of context" [4], "Smooth animated transitions may be beneficial" [6], "animation . . . can facilitate the perception of changes when transitioning between related data graphics" [33], "staged animations are recommended based on controlled experiments" [27], "visualizations . . . found to benefit from animation" [25], "use animated transitions . . . following the findings of Heer and Robertson" [32]. No mention is made in the citing works of how the generalisability of the conclusions of the original paper is limited (24 participants, two tasks, a limited range of data elements)

and the possible experimental constraints mentioned by the authors themselves (occlusion, axis rescaling, and persistent axis gridlines).[1]

In addition, the usual requirements for peer-reviewed publication of experimental studies do not typically support validation: a submitted paper described as 'only presenting incremental/already published results' tends to be immediately placed on the 'reject' pile. This situation is similar in areas of psychology [31].

Most experimental research articles or seminars conclude with a section describing the limitations of the experiment, and suggest avenues for future work. When presenting their work to an audience, it is not uncommon for a researcher, when challenged about a particular aspect of their experimental design, to respond that addressing experimental variation or validation has been 'left for future work'. Less frequent is publication of these proposed follow-on studies – if indeed they are ever conducted.

1.1   Motivation

The motivation for this paper relates to the idea of reflective critique, an idea borrowed from the practise of action learning projects. Action learning projects [13] do not aim to address specified research questions, and their approach is formative rather than summative, resulting in a cycle of continuous improvement. This approach is unlike the typical experimental research project, where a research question is clearly defined, a methodology is designed to address it, data is collected, and final conclusions made and published. In contrast, an action learning methodology encourages honest reflection on outcomes, and these reflections are fed into another cycle of the process.

This research reported in this paper arose from a reflective critique of a published graph drawing empirical study. In this case, the experimenter (the author of this paper) did not conduct the experiment and then "publish and move on", the results of the experiment having been published in a reputable journal [22]. I continued to experiment so as to explore aspects of the experimental design, to investigate subtle issues arising from the results of the first experiment, and to attempt to replicate and confirm the initial results.[2]

The reflective critique process was inspired by simple critical questions asked by audience members at two seminars which presented an experiment and its conclusions – an experiment that had already been peer-reviewed and published. The questions suggested subtle variations on the experimental method; these were not such substantial changes as to be the basis of new research questions, simply small amendments.

---

[1]I do not wish to single out this particular paper as being in any way deficient – on the contrary, it is an excellent example of a controlled information visualization experiment. Rather, I wish to prove by example my claim that authors cite the published conclusions of prior experiments, even when they have not been replicated, and when they have limited generalisability.

[2]And, needless to say, the path to having the work presented in this paper accepted for publication has not been a smooth one.

It was not necessary to investigate these issues (the paper had already been published after all), but they led me to reflect on the research, and to wish to improve the method. There was also a desire to address the audience members' observations by proving the validity of the original conclusions.

This paper therefore reports on the results of two experiments conducted 'after the fact' – after conclusions from the first experiment had been disseminated by being peer-reviewed and published [22]. Its focus is therefore two-fold: to present a series of experiments and their results, but also to describe a process whereby re-visiting empirical work has highlighted interesting facts about the process of conducting empirical studies, and, in particular, the unexpected consequence of seemingly trivial experimental decisions.

All three experiments have the same overarching research question and common aim: to determine the graph drawing layout principles favoured when participants draw graphs. They cannot be considered as three different research projects; the latter two are simply the result of reflection and critique, and were designed to address subtle issues arising from the first.

The structure of this paper is as follows: Section 2 describes the background to and motivation for the graph drawing research question, outlines the design of Experiment 1, and presents its conclusions. It then explains the issues arising from reflective critique of the experiment, and the nature of the follow-up investigations. Section 3 describes the detailed experimental processes used for all three experiments, and then describes and summarises the variations used in Experiments 2 and 3. Section 4 presents data and its analysis from Experiments 2 and 3, while Section 5 presents the results, focusing on the four key issues arising from the critique. Section 6 discusses implications of this study for both the specific graph drawing research question as well as for experimental graph drawing practise in general.

2   Experimental context

This section describes the motivation for the initial graph sketching experiment (Experiment 1) and how it relates to prior research.[3]

2.1   Research context and motivation

The domain of the experimental research is the depiction of relational data as *graphs*, where objects are represented as *nodes* (shown as small circles) and relationships between objects are represented as *edges* (shown as lines between the circles). The same relational information represented in a graph may be visually depicted in innumerable different *graph layouts*, depending on where the nodes (circles) are placed on the 2D plane and what shape the edges (lines) take as they make connections between the nodes.

---

[3]Only those aspects of the experiment relevant to this paper are described here; full details can be found in [22].

The design of automatic graph layout algorithms tend to be based on common 'aesthetic principles', for example, the elimination of edge crossings or the reduction of the number of edge bends. Early experimental research investigating the manner in which graphs may best be laid out tended to use a task-based performance approach (e.g. [11, 21, 30]), and established key findings such as the fact that a high occurrence of crossed edges reduced performance and prominent depiction of symmetry increased performance. Other experimental work considered participants' preferences for different aesthetics [19] – using the knowledge of how people would like graph drawings to be depicted in the design of layout algorithms will increase their acceptability to users of software incorporating these algorithms.

More recent empirical research has instead focused on the manner in which participants create their own visual layout of relational information as a graph drawing. Van Ham and Rogowitz [29] (abbreviated below as HR08) led the way in this approach to determining the best graph layout for human use by asking participants to manually adjust the layout of existing graph drawings. They used four graphs of 16 nodes, each with two clusters (sets of highly connected nodes). These clusters were separated by one, two, three and four edges respectively. The graphs were presented in a circular and a spring layout [8], giving a total of eight starting diagrams, shown in random order. They collected 73 unique drawings, and found that most participants separated the two clusters, that the human drawings contained 60% fewer edge crossings than the automatically produced drawings, and that humans did not value uniform edge length as much as the spring algorithm did.

Dwyer et al. [5] (abbreviated below as D+09) performed a similar hands-on experiment, asking participants to lay out two social networks, each with a circular initial arrangement. Participants were encouraged to lay the graphs out in a way that would best support the identification of cliques, chains, cut nodes and leaf nodes. With a focus on the process of layout rather than on the product, the only observation that they make about the graphs produced is that users removed edge crossings.

2.2   Experiment 1

The first experiment in the series [14, 22] was designed to address a similar research goal as HR08 and D+09, using a different methodology. The research question is *Which graph drawing aesthetics do people favour when creating their own drawings of graphs?*

There are five main design features of Experiment 1 that differentiate it from HR08 and D+09:

- The participants had to draw the graph, and then lay it out. This is a more complex task than the ones used for either HR08 or D+09;

- The participants drew the graphs from scratch, so were not biased by any initial positioning of nodes and edges (both HR08 and D+09 used an initial configuration);

Graph A	(A,D) (A,C) (B,D) (C,D) (B,C) (B.E) (C,E) (E,J) (F,G) (J,F) (F,I) (G,I) (J,H) (I,H)
Graph B	(J,F) (J,I) (G,I) (H,I) (G,H) (F,H) (G,J) (F,A) (F,G) (G,E) (A,E) (D,E) (D,C) (D,B) (C,B) (A,B) (A,C) (B,E)

Table 1: The edge lists for the two experimental graphs

- A sketching tool was used, so the physical drawing process was unhindered by an intermediate editing interface;

- Video data was collected, so both the process and product of creation could be analysed (this was done by D+09, but not HR08);

- Layout preferences were discussed with the participants in a post-experiment interview (this was done by D+09, but not HR08).

Four graphs were used, two practise graphs and two experimental graphs. Data on product, process and preferences were collected.[4]

HR08 and D+09 presented their graphs as graph drawings which already had some layout properties (circular and spring in HR08, and circular in D+09). So as not to bias the participants toward any layout aesthetics, the graphs in this experiment were presented in textual form, as a list of edges (Table 1).[5]

The conclusions of Experiment 1 were

(C1) The aesthetics that participants favoured during the process of laying out their drawings were not always evident in the final product. The videos of the layout process showed that while partial drawings clearly conformed to aesthetics (typically minimisation of crossings and grid layout) these were not evident in the final products.

(C2) Unlike most researchers in graph drawing, participants did not make a clear distinction between the creation of the graph structure and its layout – the two processes were intertwined.

(C3) The aesthetic of fixing nodes and edges to an underlying unit grid was prominent, despite prior performance experiments of graph comprehension showing no evidence of grid layout affecting performance [17].

(C4) Some demographic analysis showed a tendency for male participants with a Computing Science background to favour straight lines and a grid formation, and that these participants were more likely to recognise that a graph structure is independent of its layout.

---

[4]Note here that the publication arising from Experiment 1 also included the drawing of graphs in a point-and-click mode, but that only the sketched graph experiment is considered here.

[5]Note that in [22], this representation was referred to as an 'adjacency list', which it is not. It is simply a list of edges.

2.3   Reflective critique: issues arising

After peer-review, publication, presentation, and independent citation of this first experiment and its results, audience members at two seminars suggested some subtle variations on the experimental method: not new research questions, simply small amendments. As is typical in such situations, I responded that such variations could be addressed as part of 'future work'; however, on reflection, I decided to investigate these subtleties – to see if they would affect the results, and to test the robustness of the conclusions.

Three issues arose directly as a result of the questions posed by audience members:

- If participants compromise their layout design during creation of the drawing (C1), does this mean that they are not happy with their final product?

- If participants favour a grid-based layout (C2), would they prefer a drawing laid out using an algorithm that aligns nodes and edges to an underlying grid to their own?

- Was the tendency to favour a grid layout and straight lines (C2) a consequence of the fact that the graph information stimuli were presented as a list of edges?

2.4   Follow up investigations

Experiment 2.   Experiment 1 concluded that the aesthetics that participants favoured during the process of laying out their drawings (in particular, the grid format) were not always evident in the final product and that the participants often needed to compromise the grid structure later on in the creation process as the graph drawing became complex. Experiment 2 addressed the issue of whether participants were satisfied with their final drawing, or whether they expressed disappointment that they were unable to conform to their initial desired layout.

Experiment 1 concluded that participants favoured a grid format. A new sketch graph layout algorithm was implemented – it adapts a given sketched graph drawing so that it fixes a high proportion of the nodes and edges to a grid. Experiment 2 investigated whether the participants preferred their own sketched graph drawing, or a similar one that conformed to a grid layout.

Experiment 3.   The graph structure in both Experiment 1 and Experiment 2 was presented as a list of edges, explicitly representing each edge as a pair of alphabetic letters representing nodes. This explicit pairing of nodes may have resulted in the tendency for participants to conform to a grid structure, and, in particular, to draw edges horizontally. Experiment 3 investigated whether representing the graph structure to the participants in an alternative format also produces results that favour grid layout.
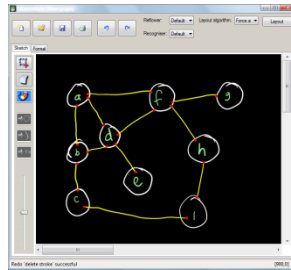
Figure 1: The SketchNode interface

3   Experimental processes

This section describes the experimental processes common to all three experiments, and then outlines how Experiments 2 and 3 differed from the initial Experiment 1. These differences are summarised in a table at the end of this section.

3.1   All experiments

The primary research question in all three of these graph sketching experiments is: *Which graph drawing aesthetics do people favour when creating their own drawings of graphs?* This is an exploratory question: participants were asked to draw graphs and their drawings were analysed for evidence of common graph drawing aesthetics.

Equipment.   A graph-drawing sketch tool, SketchNode [23] (Figure 1) was used on a tablet PC. SketchNode allows nodes, edges and node labels to be sketched with a stylus on the tablet screen, laid flat, thus allowing the same hand-movements as pen-and-paper, and allowing more natural interaction than using an editing tool. Unlike pen-and-paper, however, the SketchNode interface allows nodes (or groups of nodes) to be selected and relocated (with corresponding movement of attached edges), and nodes and edges to be erased. It thus has the advantages of sketching on pen-and-paper, as well as the editing advantages of a graph drawing tool.[6]

Task.   Participants were given a textual description of a graph on paper and asked to draw it in SketchNode, with the instruction to *Please draw this graph as best as you can so to make it "easy to understand"*. They were deliberately

---

[6]The editing process of SketchNode was adapted slightly for Experiments 2 and 3, so as to address a potential bias in Experiment 1. Previously, when a node was moved, its associated edges were straightened. As reported in [22], observation of the Experiment 1 videos revealed very few cases in which this feature affected the final graph drawings, but to ward against future possible bias, the system was changed so edges were not straightened when nodes were manually moved.

not given any further instruction as to what "*easy to understand*" means. In particular, they were not primed with any information about common graph layout aesthetics, for example, minimising edge crossings, use of straight edges etc. They were given as long as they liked to draw and adjust the layout of the graphs.

Graphs.   Two experimental graphs were used in all of the three experiments: graph A had 10 nodes and 14 edges; graph B had 10 nodes and 18 edges. These graphs were designed with the aims of HR08 in mind, both with identifiable clusters: graph A had two clusters separated by one cut edge; graph B had two clusters separated by two cut edges: Figure 2 shows these two graphs as produced by yEd's spring layout algorithm [28], clearly showing the two clusters.
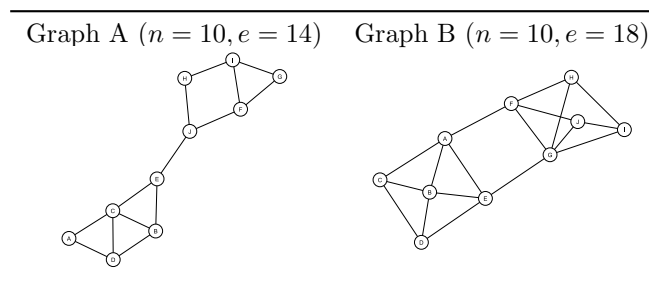
Graph A ($n = 10, e = 14$)   Graph B ($n = 10, e = 18$)

Figure 2: Graphs A and B

Experimental method.   After reading the information sheet and signing the consent form, participants answered a pre-experiment questionnaire which asked for some demographic information, including information about their experience with mathematics, theoretical computer science, graphs, and pen-based technology.

The participants were given a demonstration of the SketchNode system, including all the relevant interface features: node and edge creation using the stylus, bends and curved edges, selecting and moving nodes and edges, selecting and moving sub-graphs, labelling nodes, erasing, undoing and redoing actions, zooming and scrolling. They were also shown how to represent the textual graph information they would be given as a graph drawing, thus ensuring that they understood how to interpret the graph structure as a drawing. Participants were given ample chance to ask questions about SketchNode and the process of drawing a graph.

Besides the two experimental graphs, A and B, two practise graphs were defined, P1 ($n = 5, e = 6$) and P2 ($n = 8, e = 8$), and these were presented first. Participants were not aware that these were practise graphs – they were used to ensure that the participants were comfortable with the task and with the system before they drew the two experimental graphs. Exactly the same instructions were given to the participants for the practise graphs as for the two

	Experiment 1 ($n = 17$)	Experiment 2 ($n = 22$)	Experiment 3 ($n = 26$)
Gender	7F, 10M	7F, 12M	9F, 17M
Occupation	10 students	22 students	22 students
Age	max 49, min 20, median 24 mean 27.65	max 29, min 22, median 24 mean 24.36	max 34, min 18, median 21 mean 22
Education	13 have some university-level education	22 have some university-level education	25 have some university-level education
Experience of graphs	8 yes, 9 no	16 yes, 6 no	8 extensive, 17 moderate, 1 no
Experience of university maths/CS	8 yes, 9 no	17 yes, 5 no	10 extensive, 16 moderate
Pen-based technology experience	4 none, 9 minimal, 3 moderate, 1 extensive.	10 none, 2 minimal, 9 moderate, 1 extensive.	13 none, 11 moderate, 2 extensive.

Table 2: Demographic information about the participants for all experiments

experimental graphs: *Please draw this graph as best as you can so to make it "easy to understand".*

The two experimental graphs A and B, were then presented to the participants, with the graph edges presented in different random order for each participant. Each experiment was conducted individually, with only the experimenter and participant present. The order of presentation of the two graphs A and B was switched between participants so as to control for any possible ordering effects.

The experiments were conducted on a one-to-one basis, thus allowing observational and demographic information to be collected. The time taken for the drawing of the graphs was recorded, and at the end of the experiment, the participants were asked "Why did you arrange the graphs in the way you did?" in a recorded interview.

The participants.   Participants in all three experiments were of a similar profile: a mixture of students and non-students, of both genders, with approximately half of the student participants in each experiment studying some form of computational science. (Table 2).

3.2   Experiment 2

This experiment had three aims:

- to investigate the conclusion from Experiment 1 that *the aesthetics that participants favoured while laying out their drawings were not always evident*

in the final product (C1);

- to investigate the conclusion from Experiment 1 that *the aesthetic of fixing nodes and edges to an underlying unit grid was prominent (C3)*;

- to validate the results from Experiment 1.

Compromised layout.   The results of Experiment 1 (C1) suggests that participants might not have been entirely happy with their final drawing, as they had been obliged to compromise the favoured grid aesthetic as the graph became bigger and more complex. Experiment 2 investigated to what extent participants compromise their layout as the graph becomes more complex, being forced to abandon their desired layout aesthetics, and thus produce a drawing that they are not satisfied with.

The first research question for this experiment was "*Do participants like the layout of their final product?*". We speculated they would express dissatisfaction with their final product. Once they had drawn their graph, we asked them to indicate on a scale of 1–5 how 'happy' they were with their drawing.

Preference for a grid.   Having found that the grid layout was favoured in Experiment 1 (C3), we anticipated that participants would prefer a grid layout to their own.

The second research question was "*Do participants prefer their sketched drawing to be laid out in a grid format to their own layout?*"

The version of SketchNode used for Experiment 2 included two automatic graph layout algorithms: spring and grid. Both algorithms were specifically written for SketchNode, and importantly, they maintain the hand-drawn nature of the nodes and the edges; thus, the visual appearance of the nodes and labels remains the same, and the sketch is not 'neatened' by displaying well-formed circular nodes and straight edges. The resultant diagram is therefore clearly based on the participant's original sketched drawing, and so can be directly compared with it.

The spring algorithm was based on Fruchterman and Reingold [7]. The grid algorithm presented a hill-climbing solution whose evaluation function was based on *fundamental criteria* of no overlapping nodes, no node-edge intersection and reduction in edge crossings, and *specific criteria* of displaying the edges and nodes, as much as possible, on the lines and vertices of an underlying unit grid.

At the end of the sketching stage of the experiment, the participant's own drawing was laid out using both algorithms. Participants were asked to indicate which of the three drawings they preferred: own, grid or spring. The spring algorithm was included so as to provide a range of options: we had no particular hypothesis as to whether it would be preferred more or less than the other two options, which were of more interest.

In an attempt to eliminate any personal bias or recency effects, a willing subset of the participants also chose between hand-drawn and the two automatically-laid-out drawings two weeks after the experiment.

Validation.   As both changes to the experimental method for Experiment 2 were post-experiment activities, this experiment also served as a means to validate the results of Experiment 1.

3.3   Experiment 3

This experiment had one aim:

- to determine if there was any inherent bias in the way the abstract graph information was presented.

Effect of graph format.   The most important differentiating design feature of Experiment 1 in comparison with prior research was the manner in which the graphs were presented to the participants: HR08 and D+09 presented their graphs as concrete graph drawings which already had some layout properties (circular and spring in HR08 and circular in D+09). Experiment 1 used an abstract text representation. Here we investigated whether even this abstract form had produced a bias.

The research question for this experiment was "*Does the format in which the graph structure is represented affect the layout of the sketched graph drawings produced?*" It could have been that Experiments 1 and 2 were biased toward a grid-formation by the presentation of the graph as an edge list (as in Table 1), and that presenting the graphs in an alternative format may reveal less favouring of horizontal and vertical edges.

For Experiment 3, we presented the graphs as an adjacency list (as in Figure 3), and followed exactly the same process as Experiment 1. This format is visually quite different from an edge list, as each edge is not clearly specified as an individual pair, and it is more obvious which nodes have a higher degree. We wished to investigate whether a format that does not focus on the individual node pairs (as in Experiment 1) would result in user-sketched drawings that conform to a grid structure.

Changing the format of the stimuli for this experiment (and therefore a significant aspect of the experimental methodology) means that it could not be used to validate the results from Experiment 1 in the same way that Experiment 2 could.

3.4   Differences between the experiments

Note that because Experiment 2 and Experiment 3 focus on different aspects of the experiment (Experiment 2 focuses on the graph drawing data; Experiment 3 focuses on the graph structure stimulus), only relevant aspects of these experiments were varied.[7] Table 3 shows both the differences between the experiments,

---

[7] Of course, we could have included the post-experiment discussion questions for Experiment 2 in Experiment 3; we did not do so because we were focused on varying only one aspect of the experimental design at a time, and could not have been sure that the different methodology used in Experiment 3 would not have affected this post-experiment data collection

Graph A		
C	E B	
A	C	
F	G J	
I	H F	
B	E	
G	I	
E	J	
D	B A C	
H	J	

Graph B		
Z	Y W	
V	W P Z	
U	R T W	
T	S	
W	Y	
P	U R	
S	R P	
Y	X	
X	V W	
R	T	

Figure 3: Example adjacency lists for the experimental graphs in Experiment 3

as well as those factors that remained the same, and Figure 4 (graph A) and Figure 5 (graph B) show example sketches from all three experiments.

4 Data analysis

This section presents the raw data and the analysis of the results of Experiments 2 and 3, relating them to the four issues of compromising layout, preference for a grid, validation, and effect of graph format.

4.1 Experiment 2

Of the 44 sketch drawings produced by the 22 participants, 4 were structurally incorrect.[8] As the focus of the experiment was on how participants represented graphs (and not on whether they drew the graphs correctly or not), incorrect graphs were not removed from th analysis.

Compromising layout. We asked the participants in Experiment 2 to indicate how happy they were with their two drawings on a five-point scale (Figure 6).

Participants were asked what they didn't like about their drawings, and how they would improve them. None of them mentioned that they would have liked to conform to a grid layout, and most of the comments related to local issues like the straightening of the edges, size and shape of the nodes, connections between the nodes and edges, undesirable crossings and long edges, and straight and similar length edges. The few comments that referred to overall layout of the drawing were concerned with spreading the nodes out, symmetry and circular layout. There were also several comments about the need to plan in advance.

---

[8]27A-2 had two H nodes, 22B-2 had two U nodes, 23B-2 was missing the (V, X) edge, and 16B-2 had two Y nodes and two (X, Y) edges.

	Experiment 1	**Experiment 2**	**Experiment 3**
experimental graphs	A $(n = 10, e = 14)$ B $(n = 10, e = 18)$	as Experiment 1	as Experiment 1
experimental task	After introduction and training activities, participants sketched the two graphs	as Experiment 1	as Experiment 1
equipment	SketchNode	as Experiment 1	as Experiment 1
number of participants	17	22	26
form of graph presented to participants	(A, D) (A, C) (B, D) (C, D) (B, C) (B, E) (C, E) (E, J) (F, G) (J, F) (F, I) (G, I) (J, H) (I, H)	as Experiment 1	S | U R Q | S V | W Z Y | X V R | U W | Y U | Z T | R Q S X | Z
post-experiment discussion	none	participants indicated how happy they were with their drawing	as Experiment 1
post-experiment ranking	none	participants ranked sketched drawings against associated spring and grid drawings	as Experiment 1
later ranking (two weeks later)	none	participants ranked their own drawing against corresponding spring and grid drawings	as Experiment 1
later ranking (two weeks later)	none	participants ranked drawings from two other participants against corresponding spring and grid drawings	as Experiment 1

Table 3: Summary of the differences between the three experiments.

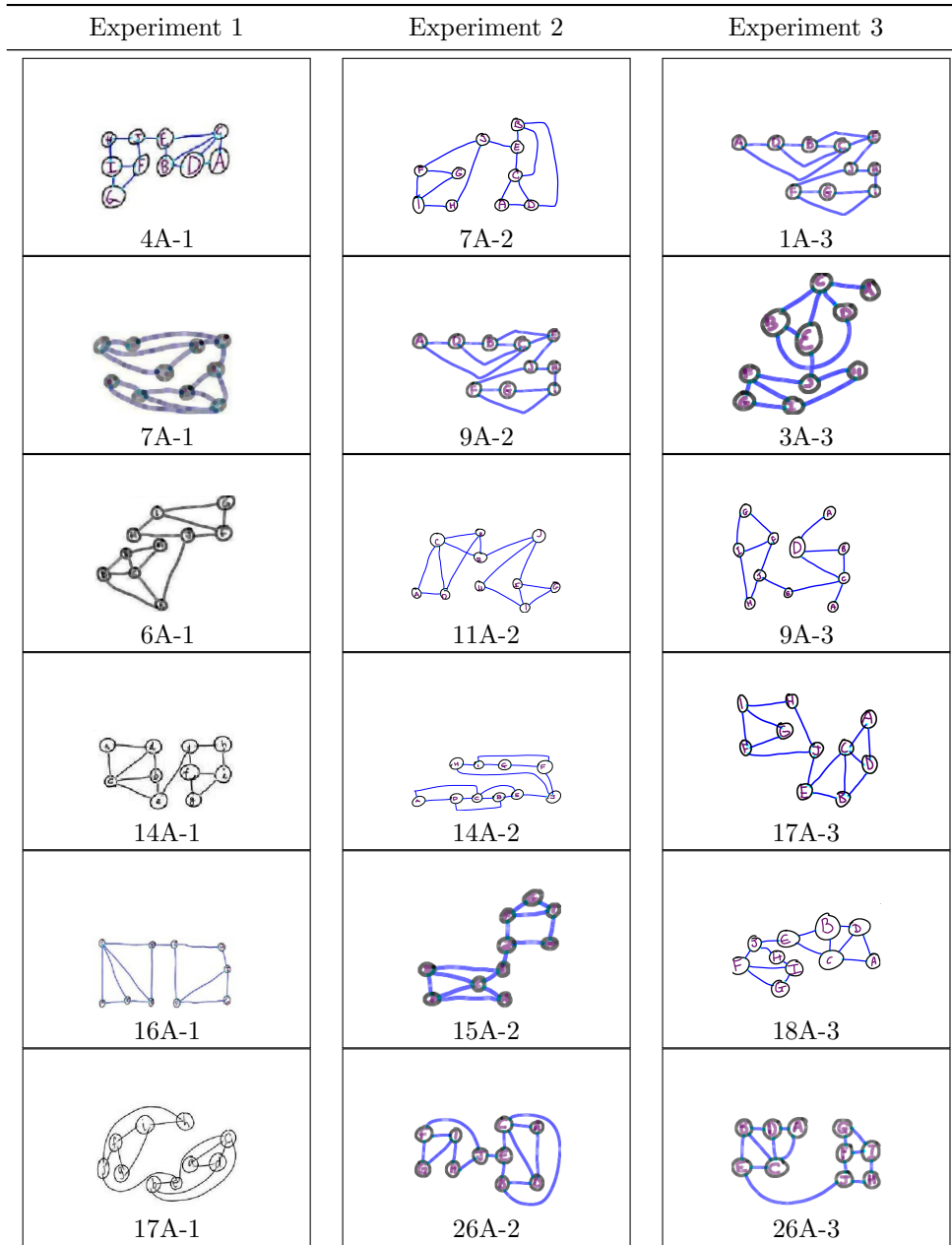| Experiment 1	Experiment 2	Experiment 3
4A-1	7A-2	1A-3
7A-1	9A-2	3A-3
6A-1	11A-2	9A-3
14A-1	14A-2	17A-3
16A-1	15A-2	18A-3
17A-1	26A-2	26A-3

Figure 4: Example drawings from all three experiments: graph A. The notational convention is: graph A drawn by participant 3 is 3A; graph B drawn by participant number 8 is 8B. The suffix -1, -2 or -3 indicates Experiment 1, 2 or 3.
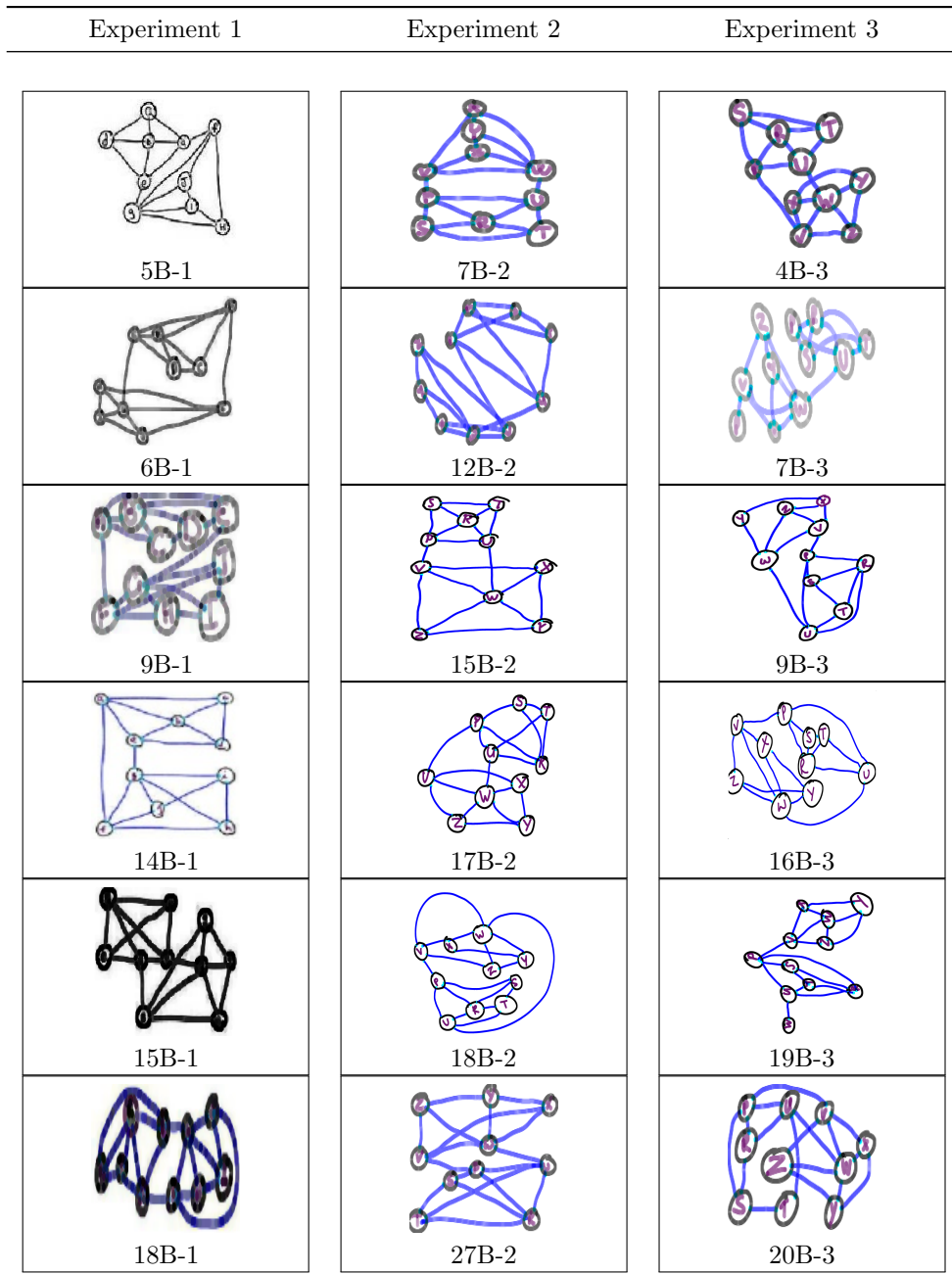
Experiment 1	Experiment 2	Experiment 3

Figure 5: Example drawings from all three experiments: graph B. The notational convention is: graph A drawn by participant 3 is 3A; graph B drawn by participant number 8 is 8B. The suffix -1, -2 or -3 indicates Experiment 1, 2 or 3.
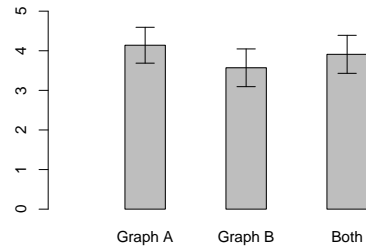
Figure 6: Participants' satisfaction with their own sketched drawing for graph A and graph B (1=not particularly happy; 5=perfectly happy). The thin black lines show the standard deviation.
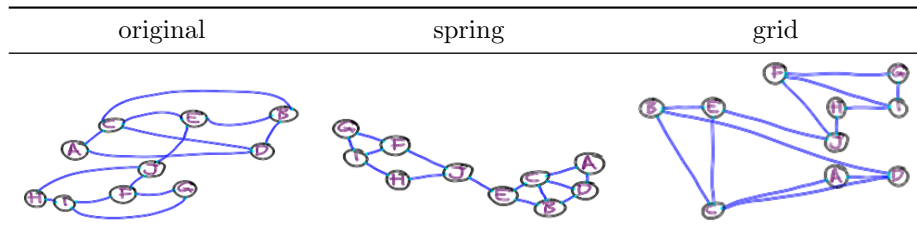

Figure 7: The TWS-GA for participant 17. Note that the hand-drawn nature and form of the nodes and their labels is retained.

| number of participants | | | mean rank: | | | % preferred | |
	/ TWSs		sketch	spring	grid	sketch	grid
After drawing the graph (own TWS)	22/22	GA	2.45	1.77	1.77	50%	23%
		GB	*2.54*	*1.41*	2.05	68%	23%
Two weeks later (own TWS)	14/14	GA	2.07	1.86	2.07	36%	29%
		GB	1.79	1.79	2.43	29%	43%
Two weeks later (own + two other TWSs)	14/42	GA	2.00	1.79	2.21	36%	36%
		GB	*1.90*	*1.60*	*2.50*	26%	55%

Table 4: Participants' ranking of TWS-GAs and TWS-GBs. GA is graph A, and GB is graph B. Significant results in *italic*.

Preference for a grid.   Once participants had sketched their own graph, the two automatic graph layout algorithms (Section 3.2 above) were applied to their drawing.

A *three-way-set* (TWS) is defined as a set of three drawings: a participant's original sketched drawing, and the two versions of this original sketch as produced by the algorithms. Each participant has their own TWS-GA and TWS-GB. Figure 7 shows a TWS for one of the participant's graphs, in its original form, and after having had the two algorithms applied.

Participants ranked the drawings in their own TWS-GA and TWS-GB at the end of Experiment 2 (Table 4 and Figure 8).

Friedman non-parametric tests with adjusted pairwise comparisons show:

- Graph A: although the rank for the participant's own graph is greater than the ranks for the spring and grid layouts, this difference is not significant ($p = 0.071$)

- Graph B: the participant's own graph has a significantly higher preference rank than the spring layout ($p < 0.001$).

After approximately two weeks, we contacted all participants for a follow-up ranking experiment; fourteen of them took part. This time, each participant was asked to rank their own TWS-GA and TWS-GB (as before), as well as the TWS-GA and TWS-GB for two other participants (those who took part in the experiment before and after them). They were not told that their own drawings were included in these sets.

There are insufficient data points for appropriate analysis on the 14 revised ranks for the participants' own drawings. In analysing the ranking for all 42 TWSs, Friedman analysis with adjusted pairwise comparisons show:

- Graph A: there is no significant difference between the ranks for the three versions of the drawings ($p = 0.145$).
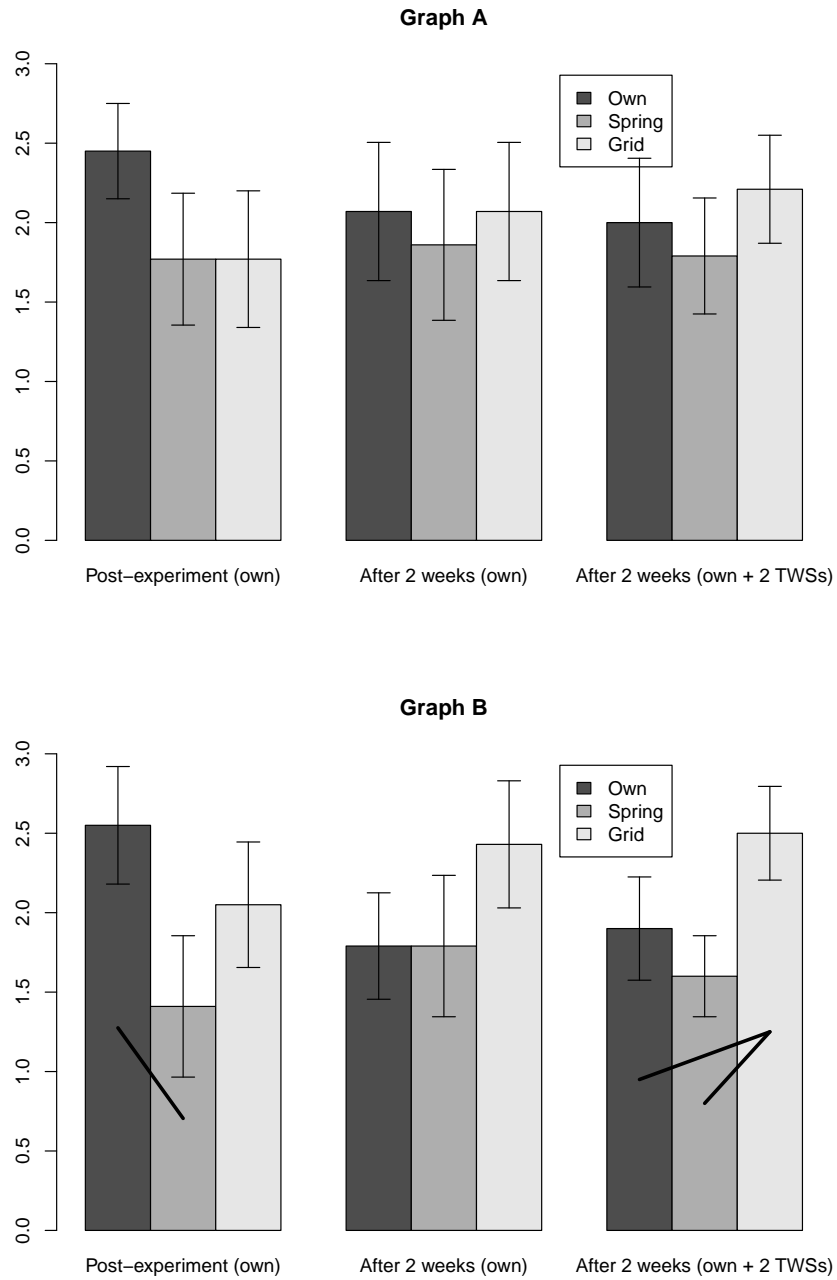
Graph A

Graph B

Figure 8: Participants' mean ranking of the TWSs. A rank of 3 represents the most preferred drawing. Standard deviations are shown. Black lines indicate significant differences.

- Graph B: the grid drawing is ranked higher than both the participants' own drawing ($p = 0.019$) and the spring layout ($p < 0.001$).

Validation.   Experiment 1 found that participants appeared to favour a grid layout and horizontal and vertical edges in their sketched graph drawings. We analysed the 44 graph drawings from Experiment 2 for the following key layout features (Table 6):

- *Number of edge crossings:* points outside of the node boundaries where one or more edges cross.

- *Number of straight lines:* Sketched drawings are unlikely to have lines that are exactly geometrically straight, so a visual assessment as to whether an edge was intended to be straight was agreed by two independent researchers.

- *Number of vertical or horizontal edges, and grid structure:* A visual assessment was agreed by two independent researchers as to whether edges were intended to be horizontal or vertical, and whether drawings had been drawn with a grid structure in mind.

To see whether the values of these metrics differed between Experiment 1 and Experiment 2, independent samples two-tailed t-tests were conducted (Table 5), using the 34 drawings from Experiment 1 and the 44 drawings from Experiment 2.

	E1 $n = 34$	E2 $n = 44$	E1 v. E2 p
Number of crosses/drawing	2.21	4.05	0.163
Number of crosses/drawing (excl. outliers)	1.24	2.05	0.166
% HV edges	33.6%	28.0%	0.222
% straight edges	88.1%	85.8%	0.638
Drawn with grid in mind	10(29%)	7(16%)	0.156

Table 5: Comparing the computational results between Experiment 1 (E1) and Experiment 2 (E2) for the purposes of validation.

4.2   Experiment 3

Of the 52 sketch drawings produced by the 26 participants, 6 were structurally incorrect.[9]   While the focus on these experiments was on how participants represented graphs (and not on whether they drew the graphs correctly or not), it is interesting to note that there are more structural errors in these graph drawings from Experiment 3 (8, that is, 0.31 per participant) than in

---

[9]5A-3 had an extra (D, F) edge, 9A-3 had 2 A nodes, and a missing (A, E) edge, 19A-3 had an extra (F, E) edge, 9B-3 was missing (W, X) and (W, V), 12B-3 had an extra (W, P) edge, and 19B-3 had two W nodes.

Experiment 1 (3, that is, 0.18 per participant) and Experiment 2 (5, that is, 0.23 per participant), even though exactly the same graphs were drawn.

Effect of graph format.   We analysed the 52 sketched graph drawings from Experiment 3 using the same features as for Experiments 1 and 2 (Table 6).

To see whether the values of these metrics differed between the two experiments, independent samples two-tailed t-tests were conducted (Table 7), using the 34 drawings from Experiment 1 and the 52 drawings from Experiment 3.

In the interviews, as before, no participants spoke directly of a grid layout; there were some comments about local features (size of the nodes, the need for straight lines), and a few about layout (crossings, symmetry, distance between nodes). When asked why they drew the graph the way they did, many said that the adjacency list itself suggested the structure of the drawing: they 'worked from top to bottom'.

Nine participants said that they could have made the drawings better if they had done more planning in advance.

5   Results

Table 8 summarises the findings from the three experiments.

The results of Experiments 2 and 3 need to be considered in the context of their aims:

- to investigate the conclusion from Experiment 1 that *the aesthetics that participants favoured while laying out their drawings were not always evident in the final product*;

- to investigate the conclusion from Experiment 1 that *the aesthetic of fixing nodes and edges to an underlying unit grid was prominent*;

- to *validate* the results from Experiment 1;

- to determine if there was any *inherent bias* in the way the graph information was presented as experimental stimuli.

Compromised layout.   We wished to see whether participants expressed any dissatisfaction with their drawings as a result of having to compromise any layout aesthetic that they may have favoured while creating the drawing. In general, the participants were satisfied with the overall layout of their own drawings: their satisfaction ratings were high, and they expressed little dissatisfaction. This was surprising, as one of the stated conclusions of Experiment 1 was that participants compromised their layout as the graphs became more complex. We thus expected that participants would be dissatisfied with their own drawings. It may be that they were not aware of any compromises they made, or that they were simply glad to have been able to fulfil the task.

	Both graphs Experiment 1	Graph A	Graph B	Both graphs Experiment 2	Graph A	Graph B	Both graphs Experiment 3
Number of drawings	34	22	22	44	26	26	52
Edge crossings							
Total number of edge crossings	74	35	143	178	30	94	124
Mean crossings per drawing	2.18	1.59	6.50	4.05	1.15	3.62	2.38
Median crossings	0.5	0	3	1	0	2	1
Max crossings	22	15	29	29	9	28	28
Drawings with zero crossings	17	14	2	16	16	6	22
Drawings with > 2 crossings	9	3	12	15	3	11	14
Total edge crossings excluding obvious outliers (8B-1, 11B-1, 24A-2, 9B-2, 20B-2, 25B-3)	42	20	70	90	30	54	84
Mean crossings per drawing excluding outliers	1.31	0.95	3.64	2.25	1.15	2.25	1.68
Vertical/horizontal edges							
Mean percentage edges horizontal or vertical	33.6%	31.5%	24.5%	28.0%	24.1%	22.6%	23.3%
Graphs clearly drawn with a grid arrangement in mind	10(29%)	3	4	7(16%)	3	1	4(8%)
Straight edges							
Mean percentage of edges straight	88.1%	87.0%	84.6%	85.8%	91.3%	89.4%	90.3%
Drawings with all edges straight	18	12	9	21	12	9	21

Table 6: Computational metrics for all the graph drawings produced for Experiments 2 and 3, with data from Experiment 1 for comparison.

	E1 $n = 34$	E3 $n = 52$	p	E1 v. E3
Number of crosses /drawing	2.21	2.38	0.829	
Number of crosses /drawing (excl. outliers)	1.24	1.62	0.440	
% HV edges	33.6%	23.3%	0.006	more HV edges in Experiment 1
% straight edges	88.1%	90.3%	0.590	
Drawn with grid in mind	10(29%)	4(8%)	0.007	more grids in Experiment 1

Table 7: Comparing the computational results between Experiment 1 (E1) and Experiment 3 (E3) for the purposes of testing the effect of graph format.

	Experiment 1	Experiment 2	Experiment 3
compromised layout	assumed, from video data	*not proven*	not investigated
preferring a grid over own drawing	assumed, from sketch graph drawing data	*some support after elimination of recency effects*	not investigated
validation	not applicable	*validation of key graph drawing metrics*	not appropriate
effect of graph format	not considered	not investigated	*effect found*

Table 8: Summary and comparison of the findings for the three experiments. Findings are shown in *italics*

This is an instance of *stating a conclusion based on speculation.* In Experiment 1, observation of the videos (by two independent researchers) suggested that participants conformed to a grid formation in the initial stages of drawing, but that this layout feature was abandoned later in the drawing process. This conclusion was, however, simply suggestive, and there was no qualitative interview data to back it up. It appears that, even if it were the case that layout compromises were made, participants were not aware of having done so.

Preference for a grid. We wished to see whether participants would prefer a grid-based layout to their own layout. While their initial responses were that their own drawings were preferred, when asked the same question without the recency effect of having just drawn the graph, the grid layout was ranked as substantially better than the sketched drawing for the more complex graph. Note that, however, none of the participants mentioned a grid formation or horizontal or vertical edges in their comments; as in Experiment 1, this feature was demonstrated, but not recognised by the participants.

It appears that participants know what they like when they see it (and when it is not in competition with a drawing that they know is their own), but cannot independently articulate the layout features that contribute to what they like.

This is an instance of *stating a conclusion based on quantified observational data.* In Experiment 1, the participants' drawings were analysed by two independent researchers who formed the same conclusion: many drawings favoured a grid layout, however, this inferred conclusion does not hold when more direct data is collected.

Validation. The opportunity to conduct exactly the same experiment more than once to verify empirical results is rare. The fact that there are no significant differences in the values for the key metrics suggests validation of the results of Experiment 1: when the graph is presented as a list of edges, participants tend to favour a grid layout and straight lines. We perhaps should not be surprised, as many of the participants for Experiment 2 fell into the category that was singled out in Experiment 1 as favouring grid and straight lines: males with a computing science or graph background.

Note that, however, none of the participants mentioned a grid formation or horizontal or vertical edges in their comments; as in Experiment 1, this feature was demonstrated, but not recognised by the participants.

As in the first experiment, participants' comments about the way in which they laid out the drawing, their criticism of their layout, and their wish that they had planned their drawing in advance, suggested that participants did not see the process of creating the graph drawings as a separate activity from laying it out; few participants drew the graph and then spent time tidying it up by re-locating nodes.

This is an instance of *validating data.* Using exactly the same experimental methodology for Experiments 1 and 2, we would expect to obtain similar values for the dependent data variables.

Effect of graph format.   We wished to see whether the form in which the graph stimuli are presented would affect the results. The drawings produced in Experiment 3 from adjacency list stimuli did not favour a grid layout; there is a significant difference on the key metrics of horizontal and vertical edges, and grid formation, between Experiments 1 and 3.

Looking at the demographic data of the three experiments might give clues as to this difference (Table 2). However, the fact that there was a preponderance of male computing science participants in Experiment 3 would suggest an even higher likelihood of grid-based sketches (C4), but this is not the case.

It appears that the format in which the relational graph information is presented to the participants affects the final layout of their drawing – a result echoing experimental results on visual metaphors [34]. The participants spoke of 'following the table from top to bottom' when creating their drawing in Experiment 3; it is likely that the participants in Experiment 1 followed the edge list from left to right – and that in both cases, the form of their sketched drawings was affected by format in which the edges were presented.

There is an irony here: one of the main reasons that Experiment 1 presented the graph structure to the participants as a list of edges was so as to address the possible layout biases in HR08 and D+09 who presented the graphs in an initial spring or circular layout. It seems that even using an edge list (as opposed to an adjacency list) can introduce a bias.

This is an instance of *unintended bias*. Even a simple (and seemingly innocuous) decision made for Experiment 1 introduced a factor that biased the results substantially.

6 Discussion

6.1 Implications for this research

Our speculation that participants would prefer the result of a grid-based algorithm over their own drawing was partially confirmed, but only when the 'personal pride' factor had diminished over time, and for the more complex graph. It is still clear, however, that while participants might prefer a grid layout in both creation and recognition, they are less able to articulate that this is a favoured attribute, and that participants do not distinguish between the two processes of creating and laying out a graph drawing.

The most surprising result is that how the graph is presented to the participants has a significant effect on the form of their drawing – this was something that had not been considered originally.

The significance of the structure of the input graph for the (human-generated) output drawing suggests the following:

- There will always be a bias relating to the manner in which information is presented to a participant;

- Even if the results of an experiment are validated by repeating it, these results may still be compromised by bias;

- For this research question, the only way that such bias may be mitigated against would be by asking the participant to draw a graph based on their internal cognitive structure, and not on an externally visible form – this may mean describing and discussing a scenario with the participant (for example, a social network) and then asking the participant to draw the graph representing the relational information. However, even a methodology like this would be susceptible to biases based on the language used to describe the scenario.

Of course, this story could not simply end here, as there are several outstanding issues that could be addressed about all three of these experiments. Do these results extend to larger graphs? What would happen if the participants were all novice computer users? Or if a digital whiteboard were used? Or if the participants were told that the graphs related to a domain (e.g. a transport network or a circuit diagram)? Or if they were to be explicitly advised to plan in advance? Further experiments would no doubt shed more light on these initial studies (and would probably reveal some unexpected results). It remains to be seen whether these 'future work' experiments will be conducted.

6.2   Implications for experimental research

We did not set out to investigate the effect of experimental subtleties: our original aim was not to run a series of comparative experiments. If it had been, then we might have conducted a broader experiments-within-an-experiment study, preferably using the same participants throughout.

We also did not employ any systematic process in collecting suitable experimental variations to investigate; the three issues arose serendipitously from audience questions – these were the only relevant experimental issues that arose on the two occasions that the work was presented.

No – we initially set out to investigate what happens when participants draw graphs from scratch – and we published a peer-reviewed paper in a reputable journal presenting the results of this study. This is common experimental practise: researchers run an experiment, collect data, form conclusions, and publish. And then typically move on to their next experiment.

The contribution of this paper is therefore broader than the simple 'run an experiment and report' model: by reflecting on and critiquing our own experimental work, and investigating issues arising from the critique, we have demonstrated the limitations of this common practise.

These conclusions can be summarised in terms of the following advice:

- to experimenters: know and acknowledge your experimental biases and limitations; pursue modest research goals.

- to researchers: only confidently cite the results of experiments that have been validated; do not generalise results outside their limitations.

Both of these points appear self-evident, and in the physical and biological sciences it would even be considered unnecessary to explicitly have to state them.

So what makes graph drawing experiments different? I suggest that there are three features of these types of experiments that make them particularly susceptible to the practises of both reporting and relying on unvalidated and potentially biased or limited results:

- Experiments in graph drawing, while using a scientific method, are a relatively new area in the long history of science, and no common replication practise has yet been established in the field. Publications replicating experiments are not valued; publications that include an honest and comprehensive list of experimental limitations are rejected as being of too narrow scope.[10]

- Human experiments (unlike many of those in the physical or biological sciences) are difficult and time consuming to run. This is not to diminish the efforts of these other sciences, simply to point out that getting representative 'samples' for such experiments (that is, human participants) is more difficult, and that the behaviour of these 'samples' when engaged in primarily cognitive tasks will always be unpredictable.

- Technology advances at a very rapid rate, which means that there is always a vast selection of new exciting experiments to conduct using new technology, seducing the experimenter away from repeating previous experiments to validate results.

It is rare that researchers repeat an experiment with subtle variations – doing so has revealed that there is still much to learn about the nature of user-sketched graphs, that even a carefully-conducted experiment may have flaws, that there is value in not simply moving on to the next 'big' question, and that repeating an experiment so as to investigate subtleties may produce surprising results. All experiments have limitations – no experiment can ever be perfect. This paper demonstrates the importance of acknowledging these limitations and of validating results where possible. In addition, published experimental results should be read with a healthy critical attitude and their findings cited appropriately, and within context.

The sketched graph data from Experiments 1, 2 and 3 can be found at `http://jgaa.info/accepted/2014/Purchase2014.18.2/SupplementaryMaterial.pdf`.

Acknowledgements

---

[10]Both these points are based on extensive personal experience.

efforts of Beryl Plimmer and Hong Yul, of the University of Auckland. The experiments were conducted as part of student projects, by Christopher Pilcher, Rosemary Baker, Anastasia Giachanou, and Gareth Renaud. Euan Freeman, Mhairi McDonald and John Hamer all assisted with various aspects of the data analysis, and John Hamer helped with paper presentation. Ethical approval was given by the University of Auckland and Glasgow University.

References

[1] D. Archambault and H. C. Purchase. Mental map preservation helps user orientation in dynamic graphs. In W. Didimo and M. Patrignani, editors, *Graph Drawing*, volume 7704 of *Lecture Notes in Computer Science*, pages 475–486. Springer Verlag, 2013. `doi:10.1007/978-3-642-36763-2_42`.

[2] J. Blythe, C. McGrath, and D. Krackhardt. The effect of graph layout on inference from social network data. In F. J. Brandenburg, editor, *Graph Drawing*, volume 1027 of *Lecture Notes in Computer Science*, pages 40–51. Springer Verlag, 1996. `doi:10.1007/BFb0021789`.

[3] H. M. Collins. The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics. *Sociology*, 9(2):205–224, 1975. `doi:10.1177/003803857500900202`.

[4] G. M. Draper and R. F. Reisenfeld. Who votes for what? A visual query language for opinion data. *IEEE Transactions in Visualization and Computer Graphics*, 14(6):1197–1204, 2008. `doi:10.1109/TVCG.2008.187`.

[5] T. Dwyer, B. Lee, D. Fisher, K. I. Quinn, P. Isenberg, G. Robertson, and C. North. A comparison of user-generated and automatic graph layouts. *IEEE Transactions in Visualization and Computer Graphics*, 15(6):961–968, 2009. `doi:10.1109/TVCG.2009.109`.

[6] J. Dykes and J. Wood. Rethinking map legends with visualization. *IEEE Transactions in Visualization and Computer Graphics*, 16(6):890–899, 2010. `doi:10.1109/TVCG.2010.191`.

[7] T. M. J. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software – Practice and Experience*, 21(11):1129–1164, 1991. `doi:10.1002/spe.4380211102`.

[8] E. R. Gansner, Y. Koren, and S. North. Graph drawing by stress majorization. In J. Pach, editor, *Graph Drawing*, volume 3383 of *Lecture Notes in Computer Science*, pages 239–250. Springer Verlag, 2005. `doi:10.1007/978-3-540-31843-9_25`.

[9] J. Heer and G. Robertson. Animated transitions in statistical data graphics. *IEEE Transactions in Visualization and Computer Graphics*, 13(6):1240–1247, Nov. 2007. URL: `http://dx.doi.org/10.1109/TVCG.2007.70539`, `doi:10.1109/TVCG.2007.70539`.

[10] K. Hornbaek. Some whys and hows of experiments in human-computer interaction. *Foundations and Trends in Human-Computer Interaction*, 5(4):299–373, June 2013. `doi:10.1561/1100000043`.

[11] W. Huang. Using eye-tracking to investigate graph layout effects. In *Proceedings of the Asia Pacific Symposium on Visualisation*, pages 97–100. IEEE, 2007. `doi:10.1109/APVIS.2007.329282`.

[12] W. Huang, S. Hong, and P. Eades. Layout effects on sociogram perception. In P. Healy and N. S. Nikolov, editors, *Graph Drawing*, volume 3843 of *Lecture Notes in Computer Science*, pages 262–273. Springer Verlag, 2006. `doi:10.1007/11618058_24`.

[13] S. Kemmis and R. McTaggart. *The Action Research Planner*. Deakin University Press, 1988.

[14] B. Plimmer, H. C. Purchase, and H. Y. Yang. SketchNode: Intelligent sketching support and formal diagramming. In *OzChi conference*, pages 136–143, 2010. `doi:10.1145/1952222.1952249`.

[15] M. Pohl, M. Schmitt, and S. Diehl. Comparing the readability of graph layouts using eyetracking and task-oriented analysis. In O. Deussen and P. Hall, editors, *Proceedings of the Fifth Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics'09, pages 49–56, Aire-la-Ville, Switzerland, Switzerland, 2009. Eurographics Association. URL: `http://dx.doi.org/10.2312/COMPAESTH/COMPAESTH09/049-056`, `doi:10.2312/COMPAESTH/COMPAESTH09/049-056`.

[16] K. Popper. *The Logic of Scientific Discovery*. Routledge, second edition, 2002.

[17] H. C. Purchase. Which aesthetic has the greatest effect on human understanding? In G. D. Battista, editor, *Graph Drawing Conference*, volume 1353 of *Lecture Notes in Computer Science*, pages 248–261. Springer Verlag, 1997. `doi:10.1007/3-540-63938-1_67`.

[18] H. C. Purchase. Sketched graph drawing: a lesson in empirical studies. In S. Wismath and A. Wolff, editors, *Graph Drawing*, volume 8242 of *Lecture Notes in Computer Science*, pages 232–242, 2013. `doi:10.1007/978-3-319-03841-4_21`.

[19] H. C. Purchase, J.-A. Allder, and D. Carrington. User preference of graph layout aesthetics: a UML study. In J. Marks, editor, *Graph Drawing*, volume 1984 of *Lecture Notes in Computer Science*, pages 5–18. Springer Verlag, 2001. `doi:10.1007/3-540-44541-2_2`.

[20] H. C. Purchase, R. F. Cohen, and M. James. Validating graph drawing aesthetics. In F. J. Brandenburg, editor, *Graph Drawing*, volume 1027 of *Lecture Notes in Computer Science*, pages 435–446. Springer Verlag, 1996. `doi:10.1007/BFb0021827`.

[21] H. C. Purchase, R. F. Cohen, and M. I. James. An experimental study of the basis for graph drawing algorithms. *ACM Journal of Experimental Algorithmics*, 2(4):1–17, 1997. `doi:10.1145/264216.264222`.

[22] H. C. Purchase, C. Pilcher, and B. Plimmer. Graph drawing aesthetics— created by users not algorithms. *IEEE Transactions in Visualization and Computer Graphics*, 18(1):81–92, 2012. `doi:10.1109/TVCG.2010.269`.

[23] H. C. Purchase, B. Plimmer, R. Baker, and C. Pilcher. Graph drawing aesthetics in user-sketched graph layouts. In *11th Australasian User Interface Conference*, volume 106 of *CPRIT*, pages 80–88, 2010.

[24] G. P. Quinn and M. J. Keough. *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, 2002.

[25] S. Rufiange and M. J. McGuffin. DiffAni: Visualizing dynamic graphs with a hybrid of difference maps and animation. *IEEE Transactions in Visualization and Computer Graphics*, 19(12):2556–2565, 2013. `doi:10.1109/TVCG.2013.149`.

[26] B. Sharif and J. Maletic. An empirical study on the comprehension of stereotyped uml class diagram layouts. In *IEEE 17th International Conference on Program Comprehension*, pages 268–272, Vancouver, BC, 2009. `doi:10.1109/ICPC.2009.5090055`.

[27] L. Shi, C. Wang, and Z. Wen. Dynamic network visualization in 1.5D. In *Pacific Visualization Symposium*, pages 179–186, 2011.

[28] yEd graph editor, 2011.

[29] F. van Ham and B. E. Rogowitz. Perceptual organisation in user-generated graph layouts. *IEEE Transactions in Visualization and Computer Graphics*, 14(6):1333–1339, 2008. `doi:10.1109/TVCG.2008.155`.

[30] C. Ware, H. Purchase, L. Colpoys, and M. McGill. Cognitive measurements of graph aesthetics. *Information Visualization*, 1(2):103–110, 2002. `doi:10.1057/palgrave.ivs.9500013`.

[31] L. Winerman. Interesting results: Can they be replicated? *Monitor, American Psychological Association*, 44(2):38, 2013. URL: `http://www.apa.org/monitor/2013/02/results.aspx`.

[32] J. Wood and J. Dykes. Visualizing the dynamics of London's bicycle hire scheme. *Cartographica*, 46(4):239–251, 2011. URL: `http://openaccess.city.ac.uk/id/eprint/538`.

[33] S. Yamaoka, L. Manovich, J. Douglass, and F. Kuester. Cultural analytics in large-scale visualization environments. *Computer*, 44(12):39–48, 2011. `doi:10.1109/MC.2011.363`.

[34] C. Ziemkiewicz and R. Kosara. The shaping of information by visual metaphors. *IEEE Transactions in Visualization and Computer Graphics*, 14(6):1269–1276, 2008. doi:10.1109/TVCG.2008.171.