Read It *This* Way: Scaffolding Comprehension for Unconventional Statistical Graphs

Amy Rae Fox[(⊠)] and James Hollan

University of California – San Diego, San Diego, CA, USA
`amyraefox@ucsd.edu`

Abstract. How do you make sense of a graph that you have never seen before? In this work, we outline the types of prior knowledge relevant when making sense of an unconventional statistical graph. After observing students reading a deceptively simple graph for time intervals, we designed four instructional scaffolds for evaluation. In a laboratory study, we found that only one scaffold (an interactive image) supported accurate interpretation for most students. Subsequent analysis of differences between two sets of materials revealed that task structure–specifically the extent to which a problem poses a mental impasse–may function as a powerful aid for comprehension. We find that prior knowledge of conventional graph types is extraordinarily difficult to overcome.

Keywords: Graph comprehension · Scaffold · Unconventional graph

1 Introduction

Imagine when reading a paper you encounter a graph, teeming with information—surely important by virtue of the precious column inches it spans. But as you scan for patterns, willing the author's insight to leap off the page, you find there is something unattainable. Like the writing of a foreign language, you see familiar symbols and structure, yet the rules for assembling these pieces into a meaningful whole are just outside your grasp. How do you make sense of the graphic?

As Larkin and Simon note, "a representation is useful only if one has the productions that can use it," [1, p. 71]. If we lack the ability to draw inferences from a representation, then we may find it largely useless. How is it that we develop such productions for *new* graphical forms, when even familiar systems (like scatter plots and line graphs) can prove challenging to interpret [2]? In this work, we build upon research on reading and graph comprehension to explore how readers make sense of an unconventional statistical graph. After generating hypotheses for instructional scaffolding techniques through observation (Study One), we evaluate their efficacy in the laboratory (Study Two). We find that even with explicit (text or image-based) instructions, the influence of prior knowledge from conventional graph forms is difficult to overcome. Our results suggest that when presenting unconventional graphical forms, effective techniques will direct readers' attention to the salient differences between their expectations and reality, and that designers mustn't take for granted that readers will notice they are dealing with an unconventional form.

1.1   Cognitive Aids for Graph Comprehension

Owing largely to their importance in STEM education, techniques for supporting graph comprehension have been a focus of research in the learning, cognitive and computer sciences. The most minimal interventions involve *graphical cues*—visual elements that guide attention, akin to gesture and pointing in conversation. Acartürk [3, 4] investigated the influence of point markers, lines and arrows on bar charts and line graphs, finding that different cues can lead readers to interpret a graph as depicting either an event or process. Similarly, Kong and Agrawala [5] proposed the term "graphical overlays" to refer to elements layered onto content to facilitate specific graph-reading tasks. Reviewing a corpus of statistical graphs in popular media they identified five common types of overlays: (1) reference structures (such as gridlines) (2) highlights, (3) redundant encodings (such as data value labels), (4) summary statistics and (5) annotations, each aimed at reducing cognitive load for particular graph-reading tasks.

Turning to more elaborate interventions, Mautone and Mayer [6] investigated techniques from reading comprehension to support meaningful processing of graphs in the college classroom. In a series of experiments, they presented learners with scatterplot and line graphs augmented by *signaling* (animations to reveal components of a graph, adding cues to highlight the relationship of depicted variables), *concrete graphic organizers* (diagrams & photographs of the real-world referents of variables in a graph) and *structural graphic organizers* (diagrams depicting a relationship analogous to the one represented in a graph). They found that the type of cognitive aids provided to learners affected subsequent structural interpretation of the graphs (measured by relational or causal statements).

Importantly however, these studies did not differentiate between prior knowledge of the *domain* and knowledge of the *graphs* [3, 4, 6]. The cognitive aids explored in this literature do not instruct users on *how* to read the graphs – the "rules" for their representational systems. Rather, it is assumed that the reader has familiarity with the type of graph being read. Scatterplots, time series and line graphs all rely on the Cartesian coordinate system, serving as a common graphical framework [7]. We are interested in what happens when presented with a graph that doesn't rely on this framework. Might we need a different type of scaffolding to learn a novel representational system?

1.2   Prior Knowledge and Graphical Sensemaking

Modern theories of graph comprehension posit a combination of bottom-up and top-down processing [8]. While the design of a graph is clearly important, so too is the nature of prior knowledge we bring to the task. When making sense of a graph, we draw on at least two sources of prior knowledge: our knowledge of the domain, and of the graphical form [2]. Scarcity from either source will impede comprehension in different ways.

Limited Prior Knowledge. If presented with an unfamiliar graph, depicting information in an unfamiliar domain, I will be unable use knowledge of one to bootstrap inferences for the other. Consider a novice physics student reading a Feynman diagram: without the requisite understanding of particle physics, they cannot reverse-engineer

the formalisms of the diagram. Without these formalisms, they cannot draw inferences about particle physics.

Limited Domain Knowledge. Alternatively, if presented with a familiar graph depicting data in an unfamiliar domain, I might draw on my knowledge of the graph system to learn something new about the content. If I know a straight line represents a linear relationship, I can infer that such a relationship exists between the (unfamiliar) variables in a line graph connected by a straight line [8]. It is this situation we aim to optimize in STEM education. Mautone and Mayer [6] demonstrated that animations, arrows, diagrams and photographs can all help students connect their prior knowledge of graphs to depicted variables, improving their ability to draw inferences about the related scientific processes. Of course, our expectations about how a graph works, if inappropriate, can also lead to systematic errors in interpretation [2].

Limited Graphical Knowledge. We are interested in the reciprocal case: an unfamiliar representation depicting information in a familiar domain. Importantly, by graphical knowledge we are *not* referring to knowledge of graphs in general–graphical competency–but rather knowledge of the rules governing a *particular* graph form. We reason that existing techniques for scaffolding are insufficient for this case, as the information added to the graphs serve only to strengthen the relationship between the graph-signs and (real-world) referents. This fails to address the learner's scarcity of knowledge for the representational system. If we cannot perform first order readings–such as extracting a data value–we cannot hope to perform second order readings–like inferring relationships between variables.

With sufficient domain knowledge, we expect that learners may be able to reverse-engineer the formalisms governing an unconventional graph. We wish to scaffold this process to support self-directed graph reading. As a first step, we select an obscure graphical form using an unconventional coordinate system so that we might shed light on the graphical framework: the foundation of the graph schema [7].

1.3   The Triangular Model of Interval Relations

Several representational systems for reasoning about time intervals have been explored in the literature [9], due largely to their importance in data analysis across the sciences and humanities. We have selected two informationally equivalent [1] types of time interval graphs, each representing the start and end time, duration, and relations between intervals.

In Fig. 1-*left*—the **Linear Model of Temporal Relations** (*hereafter* LM)—intervals are depicted as line segments along a one-dimensional timeline running from left-to-right. The left and right boundary points of a line segment indicate the start and end time, respectively, while the length of the segment indicates its duration. In the LM, the y-axis is solely exploited to differentiate between intervals, for example, by use of a label. In this way, the second dimension contains no *metric* information. As a result, intervals can be sorted along the y-axis in numerous ways (e.g. by start time, duration, alphabetically by label, etc.). As noted by Qiang et al. [10] this polymorphism prohibits the existence of a standard approach to visual pattern recognition with the LM, making it ill-suited for applications in exploratory data analysis and inspection of extremely large data sets.
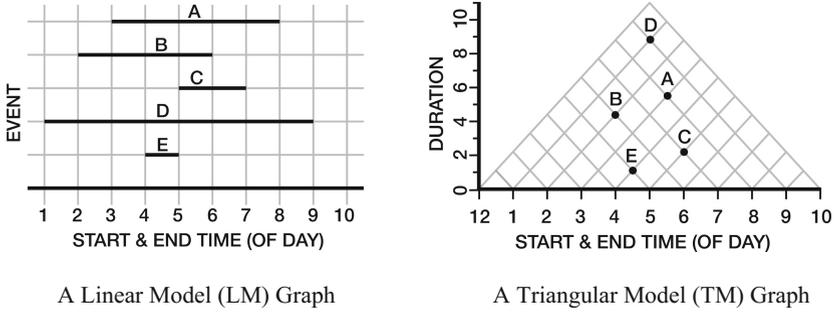
A Linear Model (LM) Graph   A Triangular Model (TM) Graph

Fig. 1. Informationally-equivalent graphs for intervals of time

Based on work by Kulpa [9] extended by Qiang et al. [10, 11] the **Triangular Model of Temporal Relations** (*hereafter* TM) overcomes this shortcoming by representing intervals as points in 2D metric space (Fig. 1-*right*). Each point represents an interval. In the vertical dimension, the height of the point indicates its duration. The intersection of the point's triangular projections (using diagonally oriented grid lines) onto the *x*-axis indicate the start and end time. In this way, every interval is represented as a unique point in the 2D graph space, and each of its elementary properties are explicitly encoded by the location of the point.

A brief inspection of the TM by even the most experienced graph readers demonstrates its relative obscurity. However, while the non-Cartesian coordinate system is unconventional, the graph depicts information about a domain in which we all share substantial prior knowledge: events in time.

1.4   The Present Studies

We are interested in what happens when experienced graph readers (undergraduate STEM majors) attempt to interpret the TM graph. Further, we wish to develop and evaluate a series of instructional scaffolds to support comprehension of the graph by self-directed readers. We start by observing students using the TM graph to solve simple questions about the properties and relations between events, and then elicit suggestions for how to make the graph easier to read (Study One). In Study Two, we evaluate four scaffolds inspired by these observations.

2   Study 1: Observing Learning of an Unconventional Graph

What strategies do we employ to make sense of an unconventional graph? In this exploratory study, we observed students solving problems with the Triangular Model (TM) graph (Part A). After a short interview, we challenged students to design instructional aids making the graph easier to read (Part B). From these data we generate hypotheses for how we might scaffold comprehension for novel statistical graphs.

2.1   Methods

Participants. Twenty-three (70% female) English speakers from the experimental-subject pool at a large American university (M(age) = 20, SD(age) = 1) participated in exchange for course-credit. All students were majors in STEM subjects. Participants were recruited in dyad pairs (9 pairs, n = 18) to encourage a naturalistic think-aloud protocol. In cases where one recruit was absent we conducted the session with the individual (n = 5), altering the procedure only by encouraging them to think-aloud as though explaining their reasoning to a partner. In total, we conducted 14 observation sessions (9 dyads, 5 individuals).

Materials and Procedure. The entire procedure ranged from 45–60 min. In Part A: The Graph Reading Task, sixteen multiple choice questions were used to probe the reader's ability to use the graph to reason about the properties of and relations between intervals. For example, a question testing the "duration" property might read: *For how many hours does event [x] last?* Participants were given one sheet of paper containing the questions and a second sheet containing a large TM graph with 15 data points[1]. After delivering instructions, we started the video recording and left the room.

Upon task completion, we conducted a short interview, prompting participants to explain how they would plot a new data point on the graph. If participants misinterpreted the graph, we began a didactic interview, prompting students to ask questions they thought might help them discover the rules of the graph system. We responded by only revealing the information explicitly requested, minimizing the effect our teaching might have on the designs produced in the next task. Once students could plot a new data point, we proceeded to Part B: The Scaffold Design Task. We asked participants to think about what they could do to make the graph easier to read for the next participant and invited them to make marks on the graph.

2.2   Study One: Results

Part A. Graph Reading Task. Participants in only 3 of the 14 sessions correctly interpreted the TM graph (M(score) = 12/16 points, (SD = 1.7), (M(time) = 19 min, SD = 30 s). These participants correctly described the graph's rules in the post-task interview. In the remaining 11 sessions, participants correctly answered only 2.2 questions on average (SD = 2.1), and were unable to correctly plot a point in the interview. Yet in these sessions, participants *did* persist in answering all questions, spending about the same amount of time on the task (M(time) = 21 min, SD = 2 min). Reviewing the artifacts participants generated gives us a window into their interpretations. Looking first at the lowest scoring sessions, we noticed many cases where participants appeared to superimpose the conventional representation for time intervals–the linear model (see Fig. 1-*left*) – atop the triangular graph (Fig. 2-*left*). We dubbed this the "linear interpretation" of the TM, which relies on participants assuming

---

[1] **Note.** All materials, data and computational notebooks for data analysis are available at https://madebyafox.github.io/Scaffolding_Graph_Comprehension

the data points are situated in a Cartesian coordinate system with a single x and y intercept. They must also infer that a point represents a *moment* in time, rather than an interval, and that the interval is represented by a line segment which they must mentally project (or physically draw) atop the graph. They must also decide which moment along an interval the point represents. In this sense, the "linear interpretation" relies on two kinds of prior knowledge: first of Cartesian coordinates in which a point has a single x-intercept, and secondly of conventions for representing intervals as linear extents, rather than points. This interpretation also requires students to ignore—or assign no meaningful referent to—the graph's diagonal gridlines. Once constructed, participants could extract information from the "linear interpretation" following the same procedure one would follow for the conventional linear (LM) graph.
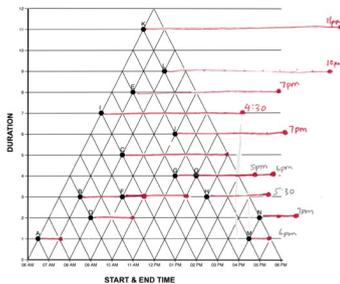
The lowest-scoring session shows an (incorrect) Cartesian interpretation.
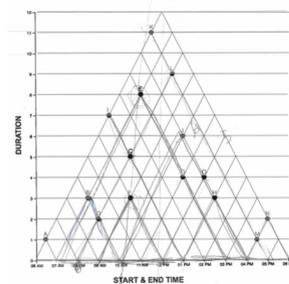
The highest-scoring session shows a (correct) triangular interpretation.

Fig. 2. Graph artifacts from lowest (*left*) and highest (*right*) scoring sessions.

Alternatively, In Fig. 2-*right* we see the artifact from the highest scoring session. Participants have reinforced the triangular intersections for several points with the x-axis. Noticeably, we do not see reinforcement of the intersections with the y-axis, presumably because this is a convention of the coordinate system participants did not need assistance to interpret.

Testing the Linear Interpretation Hypothesis. From our review of participants' graph markings, as well as the procedure they (initially) described for plotting a new data point, we hypothesized that the 11 low-scoring sessions had formed a "linear interpretation" of the graph. To test this hypothesis, we constructed an alternative answer key. First, we constructed a "linear interpretation" graph by drawing a vertical intersect for each data point to the x-axis and construing this as the *start time*. We then drew horizontal line segments from each point, with a length determined by the *duration* given on the y-axis. Using this "linear interpretation" graph, we determined the correct answer for every problem and re-scored each session. Under this alternative answer key, the mean score for the 11 lowest-scoring sessions improved from 2.2 to 8.3 (SD = 2.7 points), while the mean score for the 3 highest-scoring sessions decreased 12.3 to 3.0 (SD = 2.0 points), supporting the hypothesis that low-scoring participants interpreted the graph in accordance with the conventional linear model.

Part B. Scaffold Design Task. We evaluated the artifacts produced in response to our prompt to make the graph easier to read, and found evidence of three instructional approaches: adding pictorial intersections (Fig. 3a), providing annotations/examples (Fig. 3b, c) and text instructions (Fig. 3d).

In Fig. 3a (at right) participants have drawn attention to the diagonal gridlines and their dual-intersections with the *x*-axis by darkening and coloring them. These participants explained the most challenging part of the graph was realizing they had to look for *two* intersections with the *x*-axis.
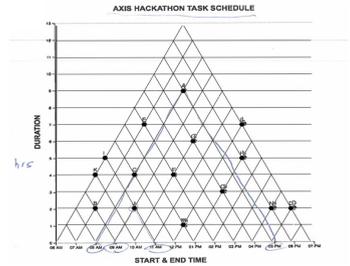
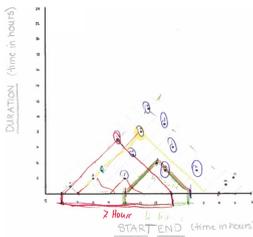**Fig. 3a.** Pictorial intersections (Color figure online)

In Fig. 3b (at left) participants have annotated their highlighted intersections. We see a partial worked example, via the annotation "7 h" to the span for the red interval.

Fig. 3b. Annotations & examples (Color figure online)

In Fig. 3c (at right) we see a worked example where participants both highlighted the intersection and gave explicit values for a sample point on the plot. Under the graph they added a production rule for finding the start-time of a hypothetical point "S", indicating that some learners may prefer text instructions. *(triangular grid faded in digital scanning)*
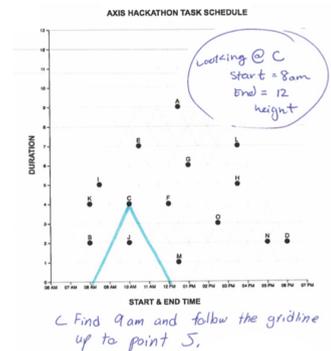

Fig. 3c. Worked example (Color figure online)

Finally in Fig. 3d (at right) we see explanatory text with an explicit definition of several graph elements.

Fig. 3d. Text instruction

2.3   Discussion of Study One

The results of Study One suggest the Triangular Model (TM) graph is challenging for STEM undergraduates. While the graph is elegant in its simplicity—as one participant noted, "once you see [the triangles], you can't (*sic*) unsee them"—most re-imagined the marks on the page as components of the more conventional representation for intervals. In interpreting this graph students invoked prior knowledge of conventions for the domain (intervals as line segments) and graphs in general (Cartesian coordinates). When prompted for instructional aids, students believed they could easily improve performance of future participants by adding instructions highlighting the multiple intersections of a point with the *x*-axis. Importantly, these scaffolds are substantively different than those explored in previous literature [2–6]. These instructions are most similar to graphical cues [3, 4], but rather than reinforcing the main argument of the graph (e.g. local maxima/minima, salient trend, etc.) they draw attention to the structure of the coordinate system. Both text and image instructions focus on the graphical framework and how to perform a first-order reading, rather than reinforcing the connection between the graph's signifiers and referents.

Owing to the limited sample size and observational methods, we fall short of explaining why some students (3 sessions) *were* able to interpret the graph while most were not. In one case, an individual interpreted the graph in the very first question, but failed to think-aloud, leaving their strategy a mystery. In the second case, the dyad pair also developed a correct model in the first question. In the third case, the dyad read the graph incorrectly for about half the questions before realizing their mistake and re-solving the problem set. These outcomes could be driven by individual differences in graphical competency, or different problem-solving strategies. Addressing this question will require further observation with directed post-task interviews.

3   Study Two: Testing Scaffolds for an Unconventional Graph

Inspired by the instructional aids produced by participants in Study One, we designed four scaffolds for self-directed learning: two text instructions (adjacent to the graphs) and two illustrations (highlighting *x*/*y* intersections). The "what-text" design (Fig. 4a) specifies the components of the graph and describes their meaning. The "how-text" design (4b) provides a set of production rules for extracting data from the graph. In the "static-image" (Fig. 4c), intersections are displayed for a single data point persistent throughout the task. In the "interactive-image" (Fig. 4d), the appropriate intersections appear & disappear when a participant hovers their mouse over any data point.

Prior work [11] has demonstrated that the computational efficiency of the TM graph can be achieved by students after 20 min of interactive video instruction. In Study Two we test the effectiveness of our designs by seeking to replicate these results with scaffolding alone. Assigning each participant to a scaffold condition, we compare their performance on both the LM and TM graphs, and subsequent ability to draw a TM graph for a small data set. We hypothesize that: (1) scaffolding will not affect performance on the LM graph, because it is conventional and relatively easy to read; (2) learners without scaffolding (control) will perform better with the LM than TM;

A point is an <u>interval of time</u>
The left intersection with the x-axis along the diagonal gridline is the <u>start time</u>
The right intersection with the x-axis along the diagonal gridline is the <u>end time</u>
The intersection with the y-axis is the <u>duration</u>.

Fig. 4a. "what-text" specifies graph components and their meaning

Fig. 4c. "static-image" displays *x/y* intersections for one data point

Start-time: follow the left-most diagonal gridline to the intersection with the x-axis
End-time: follow the right-most diagonal gridline to the intersection with the x-axis
Duration: follow the horizontal gridline to the intersection with the y-axis
Label: the letter directly above the point

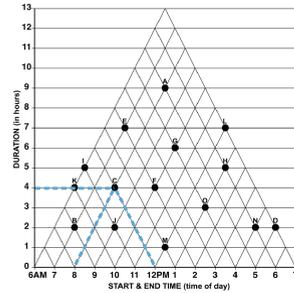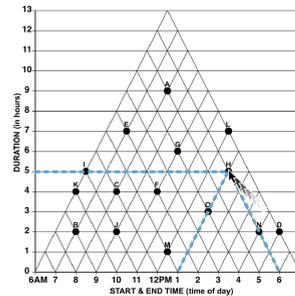**Fig. 4b.** "how-text" specifies how to extract data from the graph

Fig. 4d. "interactive-image" displays *x/y* intersections on mouseover

(3) learners with (any form of) scaffolding will perform better with the TM than LM (replication of [11]). Finally, based on observations in Study One we expect that graph-order will act as a scaffold. (4) Learners who solve problems with the LM graph *first* will perform better on the TM (relative to TM-first learners) as their attention will be drawn to the salient differences between the graph types.

3.1 Methods

Design. We employed a 5 (scaffold: none-control, what-text, how-text, static image, interactive image) × 2 (graph: LM, TM) mixed design, with scaffold as a between-subjects variable and graph as a within-subject variable. To test our hypothesis that exposure to the conventional LM acts as a scaffold for the TM, we counterbalanced the order of graph-reading tasks (order: LM-first, TM-first). For each task, we measured response accuracy and time. For the follow-up graph-drawing task, we coded the type of graph produced by each participant.

Participants. 316 (69% female) STEM undergraduates aged 17 to 33 were recruited from the experimental-subject pool at a large American university (M(age) = 21, SD (age) = 2), yielding approximately 30 participants per cell in the 5 x 2 design.

Materials

Scaffolds. For the <u>first five questions</u> of each graph-reading task, participants saw their assigned scaffold along with the designated graph. On the following ten questions, <u>the scaffold was not present</u>. Examples of each scaffold-condition for the TM graph are shown in Fig. 4. Equivalent scaffolds were displayed for the LM graph (see footnote 1).

The Graph Reading Task. Each graph reading task consisted of a graph (LM or TM) and 15 multiple choice questions (used in Study One). Questions were presented one at a time, and participants did not receive feedback as to the accuracy of their response before proceeding. The order of the first five (scaffolded) questions was the same for each participant, while the order of the remaining 10 were randomized. For each question, the participant's response accuracy (correct, incorrect) and latency (time from page-load to "submit" button press) was recorded. Because each participant completed the reading task once with each graph, we developed two matched scenarios: a project manager scheduling tasks (scenario A), and an events manager scheduling reservations (scenario B). In each scenario, an equivalent question can be identified in the other pertaining to the same interval property/relation. For example, in scenario A the question mapping to the "starts" property reads: "Which tasks are scheduled to start at 1 pm?", and the correct answer consists of 2 tasks (Fig. 5 – *left* – tasks O & H). In scenario B, the equivalent question reads: "Which reservations start at 8:00 AM?", the correct answer referencing 3 events (Fig. 5 – *right* – events D, C & L). For the LM graphs, intervals were sorted in order of duration, with the longest appearing at the top of the graph. A pilot study on Amazon Mechanical Turk using the LM graph revealed no significant differences in response accuracy or latency between the scenarios. The four graphs constructed for the study are shown in Fig. 6.

The Graph Drawing Task. Participants were given a sheet of isometric dot paper with a table of 10 time intervals, and directed to draw a triangular graph of the data ("like the triangle graph you saw in the previous task"), using the pencil, eraser and ruler provided. Isometric dot paper equally supports the construction of lines at 0, 45 and 90 degrees, minimizing any biases introduced by the paper on the features of the graph drawn by participants.

Procedure. Participants completed the study individually in a computer lab. They completed the two graph-reading tasks in sequence, one with a TM graph and the other with an LM graph (order counterbalanced). Afterward, participants completed the graph drawing task. The entire procedure ranged from 22 to 66 min.

3.2   Results: The Graph Reading Task

Performance on graph-reading tasks is a combination of response accuracy (score) and time. Table 1 displays the mean values for score (as % correct) and time (in minutes) for each graph across the scaffold conditions. As we found little variance in response

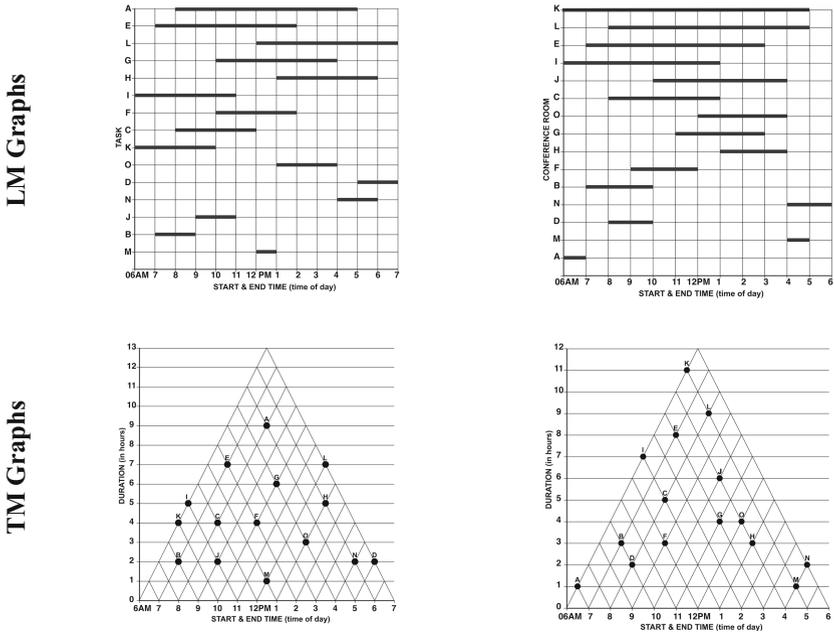**Task Scheduling (Scenario A)**   **Event Scheduling (Scenario B)**

LM Graphs

TM Graphs

Fig. 5. LM and TM graphs for each scenario of graph reading task

Table 1. Mean score and response time for graph reading tasks

	Mean score (% correct)				Mean time (mins)			
	LM graph		*TM graph*		*LM graph*		*TM graph*	
Scaffold	M	SD	M	SD	M	SD	M	SD
none-control	73	16	46	30	8.6	2.1	11.2	3.6
what-text	74	15	59	29	9.8	2.9	11.6	3.6
how-text	73	14	58	31	9.1	2.3	10.9	3.0
static-image	73	15	57	30	9.1	2.6	10.6	3.5
interactive-img	73	13	71	23	9.4	2.6	9.9	2.6
Total	73	14	59	30	9.2	2.5	10.9	3.3

time we focus our discussion on performance as judged by response accuracy. To explore the potential influence of graph, scaffold, and graph-order on scores, we performed a mixed effects ANOVA on score with graph as the within-subjects factor, and scaffold, graph-order and scenario as between-subjects factors (Fig. 6).

Effect of Graph. We found a significant main effect of graph type on score, F $(1,297) = 97.67$, p $< .001$. In Fig. 6 we see that across all factors, LM scores [green
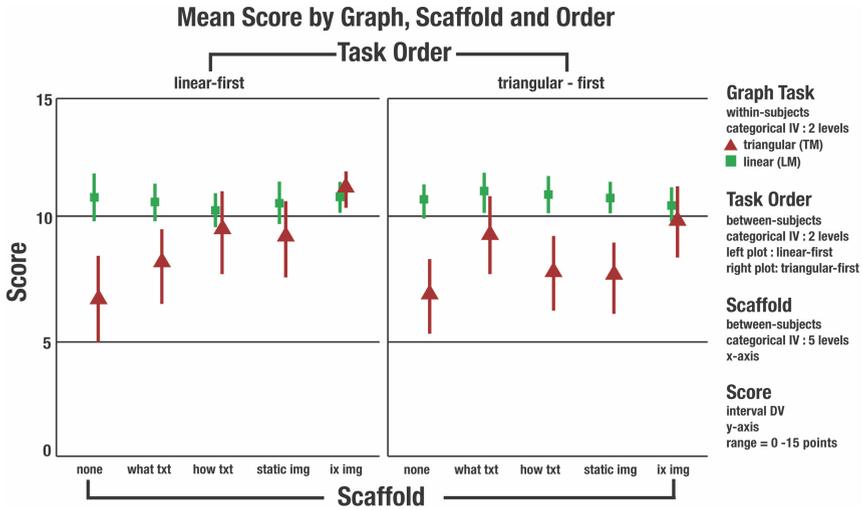
Fig. 6. Mean response score by graph, Scaffold and task order LM scores (squares) remain steady across scaffold (x-axis) and graph-order (right/left plot), while TM scores (triangles) differ by scaffold, highest in the interactive image condition. (Color figure online)

squares] (M = 10.96, SD = 2.13) are significantly higher than TM scores [red triangles] (M = 8.78 SD = 4.44), t(316) = −9.45, p < 0.001, r = 0.47, consistent with our motivating assumption that the TM graph is more challenging to interpret.

Effect of Scaffold. We found a significant main effect of scaffold on score, F(4,297) = 4.24, p < .05. A post-hoc t-test supports our second hypothesis, that across all other factors, participants in the no-scaffold control group performed significantly better with the LM graph (M = 10.98, SD = 2.33) than the TM graph (M = 6.9, SD = 4.51), t(60) = 7.07, p < 0.001, r = 0.67. Regarding our third hypothesis, we found a significant interaction between graph and scaffold, F(4,297) = 10.03, p < .001. As pre-

Fig. 7. Only the interactive image scaffold was significantly better than no-scaffold control condition.

dicted, scaffolds did *not* influence the score when solving problems with the LM (hypothesis 1), but made significant improvements in score for the TM. However, *none* of our scaffolds resulted in significantly higher scores for the TM relative to the LM. In fact, post-hoc pairwise comparisons (with Bonferroni correction) on TM scores showed
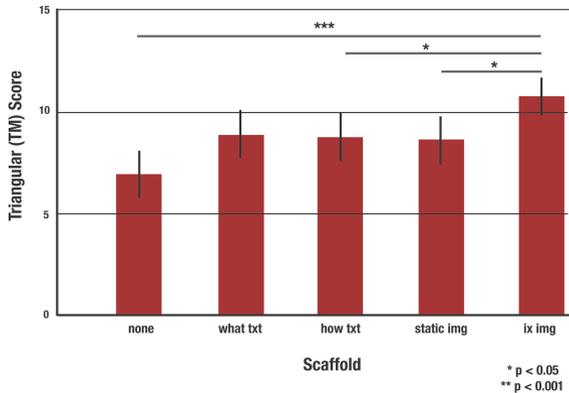
that *only* the interactive image scaffold yielded scores significantly higher than the no-scaffold control (Fig. 7).

Effect of Graph-Order. Counter to hypothesis 4 that graph-order would act as a scaffold for comprehension, we found no main or interaction effects for graph-order on response accuracy. Perhaps in order to glean salient differences between the TM and LM graphs, they need to presented simultaneously (as in Fig. 1).

Effect of Scenario. As our mixed design necessitated the use of two matched scenarios, we tested for effects of scenario in our statistical model. Unexpectedly, we found a main effect of scenario on score, $F(1,297) = 22.29$, $p < .001$, and significant interaction between graph and scenario, $F(1,297) = 34.34$, $p < .001$. When answering questions in the "task scheduling" scenario A (M = 9.20, SD = 4.12), participants had significantly lower scores, $t(316) = -4.77$, $p < 0.001$, $r = -.26$, compared to the "events scheduling" scenario B (M = 10.52, SD = 2.97). In an online pilot we found no significant differences in performance between the scenarios when tested with the LM graph. To explore the source of this effect, we examined the data sets constructed for each scenario, and in particular, the very first question students solved with the TM graph. In the "task scheduling" scenario A (Fig. 8–*left*) we see that if a learner makes the most common mistake—seeking an orthogonal intersection from the *x*-axis—there is a single data point that intersects the line: an available answer. However, in the "events scheduling" scenario B (Fig. 8—*right*), there is **no** intersecting data point. Students who were randomly assigned to this second scenario received implicit feedback that they were misreading the graph if they sought the orthogonal intersect because there was no answer to the question. We suspect this drove students to re-evaluate their strategy, yielding significantly higher scores for the "events scheduling" scenario.
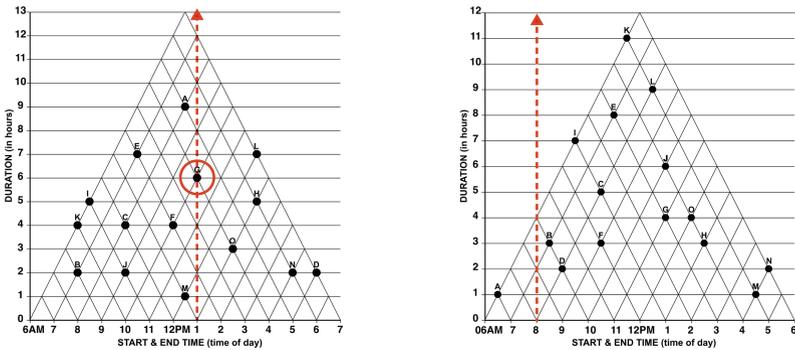
What tasks start at 1pm? A data point intersects the erroneous orthogonal projection.

What events start at 8 am? No data points intersect the mistaken erroneous projection.

Fig. 8. First question for the task (left) and event (right) scenarios.

3.3   Results: The Graph Drawing Task

The graph drawing tasks allows us how to explore how each scaffold supports students learning the graphical framework of the TM. We expect that accurately drawing requires deeper understanding of how the graph works, and analysis of any systematic mistakes students make in drawing may reveal sources of difficulty in comprehension. Following the directed approach to qualitative content analysis [12], a team of 3 raters classified all 316 drawings first into *a priori* categories [triangular, linear, other] and finally into 5 categories based on the data present in the sample: (correct) triangular, linear, scatterplot, "asymmetric triangular" and "right-angled". Interrater reliability was high ($\alpha = 0.96$) and disagreements were resolved through negotiation. The majority (73%) of participants drew correct TM graphs. 17 individuals (5%) constructed LM graphs, while 3 participants drew scatterplots with start & end time on the *x/y* axes respectively. Most interesting were the two alternative triangular forms constructed by 66 (21%) individuals: right-angle triangle, and asymmetric triangles (described in Fig. 9).

While the overall distribution of graph drawing-types was too heterogeneous to reliable correlate with TM task performance, we did examine the performance of the subset of participants who produced the two alternative triangular forms. TM scores for participants who drew "right-angle" graphs were significantly lower (M = 2.3, SD = 1.98) than for participants who drew "asymmetric triangle" graphs (M = 8.55, SD = 3.73), t(27.11) = –7.36, p < 0.001, r = 0.82.
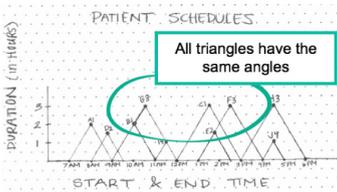

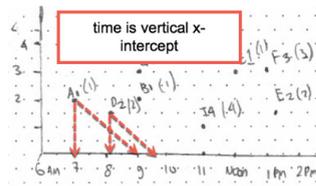
Fig. 9a. 230 students drew correct TM graphs

Fig. 9c. 44 students drew "right-angle" graphs. They plot duration on the Y axis and the interval as a point, but mistakenly use an orthogonal x-intersect for start time
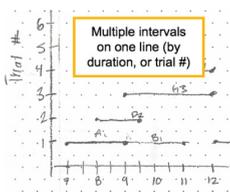

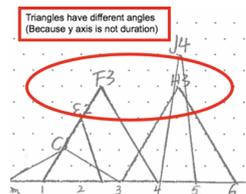**Fig. 9b.** 17 students drew LM graphs

Fig. 9d. 22 students plotted the vertical intersection as the midpoint of the interval, but the triangles were not geometrically similar because duration was not on the y-axis.

3.4 Discussion of Study Two

The results of Study Two leave us with a conundrum: why were the scaffolds designed by learners in Study One largely ineffective?

None of our designs replicated the results of Qiang et al. [11] which yielded better performance with the TM than LM graph, though there were notable differences in our tasks, including their use of an interactive graph interface with hundreds of data points, and feedback in the video instruction. Setting aside the differences in performance *between* the LM and TM graphs, we assessed the efficacy of scaffold designs by looking at TM scores alone. The widely-held assertion of Study One participants that simple text and image instructions would dramatically improve readability of the graph were not borne out, as on average, participants who received the static scaffolds performed no better than those who received (as participants in Study One) no graph instructions at all (Fig. 7).

We suspect the source of this discrepancy lies in a hindsight bias. Once students understand how the graph works, they cannot "unsee" it, and therefore underestimate the strength of their prior expectations. The unexpected effect of scenario on TM scores supports this interpretation, as students who received implicit feedback they were reading the graph incorrectly (because there was no available answer) performed better than those who did not (Fig. 8 right vs. left). In this way, the structure of the task presented the reader with a *mental impasse* [13] where their expectations (based on prior knowledge of Cartesian graph forms) left them with no solution, and their attention was actively redirected to reconsidering these expectations. The role of attention can also address why the interactive image was superior to the static text and image scaffolds. If it is the case that a reader does not *realize* they are misreading the graph (as we observed in Study One), it is easy to ignore the static scaffolds. However, it is much more difficult to ignore a stimulus that appears every time the mouse is moved over a data point. To critically evaluate the role of attention in our ongoing studies we are employing both mouse and gaze-tracking to quantify the extent and time-course of attention paid to both scaffolds and graph features.

As in Study One, the most substantial open question in this work remains the source of individual differences. Across all conditions, we see a high standard deviation (30% or 5 points) in score, again with some participants in the no-scaffold control able to correctly interpret the graph. In our ongoing work we seek to address this question with post-task interviews that prompt participants to explain their interpretation strategy while viewing a screencast replay of the their problem-solving session.

4 General Discussion

While the Triangular Model (TM) graph is elegant in its simplicity, the results of our studies demonstrate this simplicity is *deceptive*. Without assistance, most readers misinterpret the graph as the conventional representation for time intervals: the linear model. Even with cognitive aids, many students persisted in this erroneous interpretation, and only an interactive image scaffold significantly improved comprehension.

These results have implications for both the design of scaffolds and of unconventional graphs. First, when designing scaffolds one should consider the reader's expectations based on the conventional representation for variables in the domain. It is from that prior knowledge that readers begin their interpretation, *not* from a blank-slate (i.e. general graph schema) we might expect based on a graph's surface features. To overcome this, our results suggest that techniques actively directing attention to salient differences may prove most effective. The interactive-image scaffold achieves this through repeated, user-driven exposure to the multiple intersections of a TM data point with the *x*-axis. Similarly, the *mental impasse* provided by the questions in our event-scheduling scenario actively directed readers' attention to their mistaken interpretation. We are presently conducting a follow-up study testing the relative efficacy of attention-directing explicit (e.g. interactive image) and implicit (e.g. mental impasse) scaffolds.

When constructing unconventional graphs, a designer's priority is the computational affordances making the new graph-form suitable to the data and task. But as we learn from these studies, a designer should also ask, "What expectations will be invoked by the marks on the page?" For the TM graph, we suspect it is the orthogonal axes that drive readers to expect a single orthogonal intersection for each data point. But there is—strictly speaking—no reason that the axes *need* to be orthogonal. In fact, one clever participant in our graph drawing task produced what we believe to be a substantial improvement upon the TM graph, where the *y* axis was positioned diagonally on the left side of the graph's "bounding triangle". We are presently conducting a follow-up study to investigate alternative axis and grid designs, hypothesizing that such diagonally positioned axes will yield significantly better performance.

In this work, we have explored only a small subsection of the total design space of scaffolding techniques for a particular kind of unconventional graph. We expect our conclusions generalize to unconventional coordinate systems, but that other techniques need to be explored when employing unconventional *markings*. Our choice of scaffolds was inspired by direct observation and participatory design, however, we suspect a wider range of techniques might be effective in more instructional settings, including explication of worked examples, or seeing the graph being *drawn*. While we chose to separate our text and image scaffolds to test their differential efficacy, a combined text/image annotation could prove effective even in static media, and is a part of our ongoing work.

We started by reasoning that existing scaffolding techniques would be insufficient for unconventional graphs because learners would lack the prior knowledge of the new graph system required to make use of them. As Pinker [7] suggests, when confronted with an unfamiliar graph form, the reader instantiates a generic "general graph schema". However, it seems that despite differences in surface structure, a learner's prior knowledge of *other* graph forms can actively interfere with interpretation of a new graph. The novelty of the diagonal gridlines in the TM graph was not enough for most learners to suspend their Cartesian expectations. To overcome this prior knowledge, we think that successful scaffolds for unconventional graphs must not only show or tell us how to read them, but to rather alert us that that we need to pay attention, and reconsider our expectations in the first place.

References

1. Larkin, J., Simon, H.: Why a diagram is (sometimes) worth ten thousand words. Cognit. Sci. **99**, 65–99 (1987)
2. Shah, P., Hoeffner, J.: Review of graph comprehension research: implications for instruction. Educ. Psychol. Rev. **14**(1), 47–69 (2002)
3. Acartürk, C.: Points, lines and arrows in statistical graphs. In: Cox, P., Plimmer, B., Rodgers, P. (eds.) Diagrams 2012. LNCS (LNAI), vol. 7352, pp. 95–101. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31223-6_13
4. Acartürk, C.: Towards a systematic understanding of graphical cues in communication through statistical graphs. J. Vis. Lang. Comput. **25**(2), 76–88 (2014)
5. Kong, N., Agrawala, M.: Graphical overlays: using layered elements to aid chart reading. IEEE Trans. Vis. Comput. Graph. **18**(12), 2631–2638 (2012)
6. Mautone, P.D., Mayer, R.E.: Cognitive aids for guiding graph comprehension. J. Educ. Psychol. **99**(3), 640–652 (2007)
7. Pinker, S.: Theory of graph comprehension. In: Freedle, R. (ed.) Artificial Intelligence and the Future of Testing, pp. 73–126. Erlbaum, Hillsdale (1990)
8. Shah, P., Freedman, E.G., Vekiri, I.: The comprehension of quantitative information in graphical displays. In: Shah, P., Miyake, A. (eds.) Cambridge Handbook of Visuospatial Thinking (2005)
9. Kulpa, Z.: A diagrammatic approach to investigate interval relations. J. Vis. Lang. Comput. **17**(5), 466–502 (2006)
10. Qiang, Y., Delafontaine, M., Versichele, M., De Maeyer, P., Van de Weghe, N.: Interactive analysis of time intervals in a two-dimensional space. Inf. Vis. **11**(4), 255–272 (2012)
11. Qiang, Y., Valcke, M., De Maeyer, P., Van de Weghe, N.: Representing time intervals in a two-dimensional space: an empirical study. J. Vis. Lang. Comput. **25**(4), 466–480 (2014)
12. Hsieh, H.-F., Shannon, S.E.: Three approaches to qualitative content analysis. Qual. Health Res. **15**(9), 1277–1288 (2005)
13. Ohlsson, S.: Information-processing explanations of insight and related phenomena. In: Advances in the Psychology of Thinking, pp. 1–44 (1992)