Benefits of Graphical and Symbolic Representations for Learning and Transfer of Statistical Concepts

David W. Braithwaite (dwbraith@indiana.edu)
Robert L. Goldstone (rgoldsto@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E. 10th Street
Bloomington, IN 47405 USA

Abstract

Past research suggests that spatial configurations play an important role in graph comprehension. The present study investigates consequences of this fact for the relative utility of graphs and tables for interpreting data. Participants judged presence or absence of various statistical effects in simulated datasets presented in various formats. For the statistical effects introduced earlier in the study, performance was better with graphs than with tables, while for the effect introduced last in the study, this trend reversed. Additionally, in the later sections of the study, responses with graphs, but not tables, reflected increasing influence from the presence of stimulus features which had been relevant earlier in the study, but were no longer relevant. The findings suggest that graphs, relative to tables, may better facilitate perception of complex relationships among data points, but may also bias readers more strongly to favor some perspectives over others when interpreting data.

Keywords: representations; graphs; tables; mathematics; statistics; human factors

Introduction

Humans have devised a variety of different formats for externally representing information. Often, the same information may be represented in multiple representations that are *informationally equivalent*, in that each may be reconstructed perfectly on the basis of any other. Despite such equivalence, different representations may support performance of specific cognitive tasks at different levels of efficiency. Such differences have important implications for the selection and design of external representations.

The present study explores such differences with respect to graphs and tables, two of the most commonly-employed representational formats for quantitative information in a variety of fields. The relative advantages of graphs and tables have been the subject of extensive research. Tables appear to be at least as effective as graphs with respect to point reading tasks, which require one to estimate or read off individual data points (Meyer, Shamo, & Gopher, 1999; Porat, Oron-Gilad, & Meyer, 2009; Vessey & Galletta, 1991). However, graphs have often shown advantages for tasks involving complex relationships between multiple data points, such as estimating or comparing differences between points (Schonlau & Peters, 2012; Vessey & Galletta, 1991), projecting trends (Meyer et al., 1999), and detecting changes in function parameters (Porat et al., 2009).

Models of graph comprehension (Carpenter & Shah, 1998; Ratwani, Trafton, & Boehm-Davis, 2008) suggest a possible explanation for the latter findings. According to these models, spatial configurations of data points are the raw material on which graph comprehension processes operate. Importantly, some configurations may be directly perceived as basic visual features (Pomerantz & Portillo, 2012), allowing relationships between points to be "read off" directly without first encoding each point separately (Carpenter & Shah, 1998; Porat et al., 2009). For example, distances between points may be used to determine or estimate differences in the values they represent, without the need to encode those individual values at all (Pinker, 1990). Thus, in graphs, spatial configurations can act as cues for recognizing relationships between data points. Because such cues are unavailable or less salient in tables, this property of graphs can account for their observed advantages in conveying relationships among data points.

Many studies comparing task performance with graphs and tables have employed univariate datasets (Meyer et al., 1999; Porat et al., 2009). Consideration of bivariate data introduces another difference between graphs and tables. In graphs of bivariate data, there is a representational asymmetry between the two independent variables, in that one is often laid out along a spatial axis, typically the x-axis, while the other is typically represented by a non-spatial visual feature such as line color or thickness. For tables, on the other hand, such representational asymmetry is reduced, because the levels of both independent variables are laid out along spatial axes, albeit horizontal in one case and vertical in the other.

Can such representational asymmetries as exist in graphs of bivariate data lead to performance asymmetries in tasks involving one or the other variable? A few studies have provided evidence in the affirmative (Carpenter & Shah, 1998; Shah & Freedman, 2011). For example, Shah and Freedman (2011) found that when asked to interpret graphs of bivariate data, participants were more likely to describe main effects of the variable depicted in the legend than of that depicted on the x-axis, and were more likely to describe interaction effects as moderating effects of the legend variable on the effect of the x-axis variable than vice versa.

Such *representational* asymmetries in graphs, together with the intuition that these asymmetries are reduced in tables, suggest that *performance* asymmetries between tasks relating to one or the other independent variable in bivariate data should be greater for graphs than for tables. While a few studies have compared performance with graphs and tables on tasks involving bivariate data (Schonlau & Peters, 2012; Vessey & Galletta, 1991), the specific issue of how display format affects performance asymmetry between tasks has not been directly investigated.

Consideration of multiple tasks introduces the possibility of transfer, in which experience with one task affects subsequent performance on other tasks. Such transfer could be positive or negative, depending on whether previously-learned skills are correctly adapted for novel tasks, or applied without adaptation despite being inappropriate. Differences in the methods used to comprehend data in different formats, such as greater reliance on spatial configurations in graphs than in tables, could cause differences in ease of adaptation to novel tasks. Consistent with this possibility, Porat et al. (2009) found evidence of greater negative transfer between tasks for tables than for graphs of univariate data. However, it is unclear whether, and to what extent, these findings may generalize to other tasks, and in particular, to tasks involving bivariate data.

A related issue is how best to promote future positive transfer, and reduce negative transfer, when instructing learners to perform particular tasks. Educational theories (e.g. Ainsworth, 2006) suggest that incorporating multiple representations into instruction may be one path to these goals. Learners who integrate knowledge from multiple representations to form unified internal concepts are likely to show more robust and flexible learning. Analogy research suggests that comparison is a powerful tool to facilitate such integration and thus encourage positive transfer. For example, Gentner, Loewenstein, and Thompson (2003) found that management students who compared case studies illustrating a negotiation technique were more likely to apply the technique to novel cases. Considering these two lines of research together suggests that comparing graphs and tables illustrating a concept may encourage learners to learn the concept in a more abstract way, and thus to apply and adapt them more flexibly when faced with novel tasks.

The preceding discussion suggests several questions, which were investigated in the present study. First, for tasks focusing on one or the other variable in bivariate datasets, does graphical presentation lead to greater performance asymmetry than tabular presentation with respect to the depicted variables? Second, do graphs or tables show more positive (or less negative) influence of previous task practice on novel task performance? Third, does comparing graphs and tables during training promote such positive transfer (and/or reduce negative transfer)?

Method

Participants received tutorials on different types of statistical effects in the context of 2×2 factorial designs with one experimentally-manipulated variable, or "treatment factor," and one observed variable, or "secondary factor." The first two tutorials involved, respectively, main effects of the treatment factor and interaction effects of the two factors. Each tutorial explained how to judge the presence of the given effect in graphs and tables. Each tutorial was followed by a test requiring participants to judge whether the given effect was present in a series of graphs and tables.

The first two tutorials and tests were followed by a third tutorial and test pertaining to main effects of the secondary factor. This test required participants to perform the same task as for main effects of the treatment factor, namely marginalizing over one of the two factors, and differed only in which factor was to be marginalized. Comparing performance across test sections allowed us to tell whether the size of performance asymmetries across tasks differed by representational format. Further, the first two tutorials explained explicitly how to determine whether the given effects were present. By contrast, the third tutorial, regarding main effects of the secondary factor, did not. Thus, the third test provided a measure of transfer to a novel task following practice with other related tasks. The tests following each tutorial also included stimuli in a verbal format which was not shown during training. Performance with these stimuli served as a measure of knowledge transfer to a task involving a novel representation.

Participants

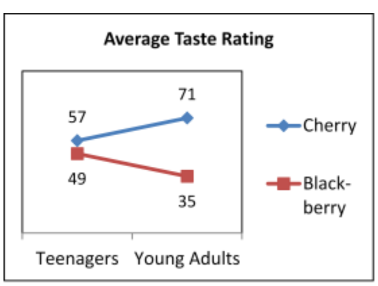
Participants were N=127 undergraduate students from the Indiana University Psychology Department who participated in partial fulfillment of a course requirement.

Materials

A set of tables, graphs, and text passages representing possible outcomes of a fictional study were developed for use as test stimuli (Figure 1). The study involved a drink taste test with two binary independent variables, drink flavor and participant age group, and one continuous dependent variable, taste rating. Drink flavor is referred to as the "treatment factor," and age group as the "secondary factor."

Each stimulus represented a dataset comprising one taste

(a)



(b)

Average Taste Rating

	Teenagers	Young Adults
Cherry	57	71
Blackberry	49	35

(c)

Among Teenagers,
the average taste ratings were
57 for cherry and 49 for blackberry.

Among Young Adults,
the average taste ratings were
71 for cherry and 35 for blackberry.

Figure 1. Test stimuli in (a) graph, (b) table, and (c) verbal format for a single dataset.
The pictured dataset shows a treatment effect and a treatment × secondary interaction, but no secondary effect.

rating for each combination of factor levels. 2 datasets were generated for each combination of presence or absence of effects of the treatment factor, secondary factor, and their interaction, yielding 16 datasets. Each effect appeared in exactly half of the datasets, and no effect was correlated with any other. 3 stimuli were created for each dataset by presenting the data in each of 3 formats: table (Figure 1a), graph (Figure 1b), and verbal (Figure 1c), yielding 48 stimuli. The secondary factor always appeared on the horizontal axis of the graphs and tables, while the treatment factor was laid out vertically in the tables and the graph legends, but these orientations were reversed in the verbal stimuli.

Another fictional study, involving effects of cognitive enhancement drugs on test scores of males and females, was devised as a basis for examples to be shown in the tutorials. Analogous to our terminology for the test stimuli, drug is referred to as the treatment factor and sex as the secondary factor. As for the test stimuli, 3 effects of these factors were possible: treatment effect, secondary effect, and treatment × secondary interaction. For each effect, 2 datasets were developed: a "positive" dataset, which had the effect, and a "negative" one, which did not. Using the same conventions as for the test stimuli, one graph and one table were created for each dataset, yielding 4 examples for each effect.

Procedure

The experiment was divided into 3 sections, one for each effect. Each section consisted of a tutorial, followed by a test, for the given effect. The sections were always presented in the same order, namely (1) treatment effect, (2) interaction effect, and (3) secondary effect. The tutorials and tests were presented via a computer interface.

The tutorials for treatment and interaction effects followed the same structure. First, participants were shown a brief description of the cognitive enhancer study, together with 2 of the 4 examples for the given effect shown side-by-side, and asked to judge whether or not the examples showed the given effect. Second, they were told that the presence of the effect depended on certain values, i.e. difference in drug scores when marginalizing over sex in the case of treatment effect, or difference of differences between drugs for each sex in the case of interaction effect. They were required to calculate and compare the relevant values, and were then told in which example(s) the effect was present, using the calculated values as justification[1]. Next, participants were asked to compare the two examples. Finally, the above procedure was repeated for the remaining 2 examples for the given effect.

The tutorial for secondary effects followed the same pattern as those for treatment and interaction effects, except that participants were not told which values they should calculate in order to judge the presence of secondary effects. Instead, after selecting which of the example(s) they thought showed effects of the secondary factor, they were asked to state how they thought the judgment should be made. They were given no feedback on their responses to this question.

Each participant was assigned randomly to one of three training conditions, which determined which examples were shown together in the tutorials. (1) In the Comparing Representations condition, the two positive examples, i.e. one graph and one table, were shown together first, followed by the two negative examples, again one graph and one table. (2) In the Contrasting Examples condition, the positive and negative examples in table format were shown together first, followed by the positive and negative examples in graph format. (3) In the Control condition, the positive table and negative graph examples were shown together first, followed by the negative table and positive graph examples.

The Comparing Representations condition directly implemented the idea, described in the introduction, of encouraging learners to compare different representations of the same information. The Contrasting Examples condition was intended as a pedagogically plausible alternative approach that employed the same materials, and involved the same amount of training, but did not afford the above opportunity for comparison of different representations. The Control condition was intended as a baseline with the same materials and same amount of variation across examples as the other two conditions, but with the examples paired in a way not expected to be useful for learners. N=42 participants were assigned to Comparing Representations, N=41 to Contrasting Examples, and N=44 to Control.

Each tutorial was followed by a test. Participants were shown a description of the taste test study and told that they would need to judge whether or not the effect about which they had just learned was present for various outcomes of the study. For each trial, one test stimulus appeared and remained onscreen until a response was received. No feedback was given. Each test stimulus was presented once per test section, in random order, for a total of 48 trials.

The experiment may be viewed online at http://perceptsconcepts.psych.indiana.edu/experiments/dwb/MRIS_02/experiment_demo_live.html.

Results

Mean accuracy on test trials was 66%, and ranged from 25% to 100%. Accuracy was significantly higher than chance, i.e. 50%, for all test sections and stimulus formats.

Accuracy scores were submitted to a 3×3×3 mixed ANOVA with training condition as a between-subjects factor, and test section and stimulus format as within-subjects factors. The main effect of training was not significant, $F(2,124)=1.82$, $p=.166$, nor were any of its interactions with other factors. The main effect of section was significant, $F(2,248)=23.67$, $p\approx.000$, indicating that accuracy was highest in the treatment section (74%), lower in the interaction section (69%), and lowest in the secondary section (63%). There was a marginal main effect of format, $F(2,248)=2.82$, $p=.061$, qualified by a significant section × format interaction, $F(4,496)=11.54$, $p\approx.000$. Accuracy scores by test and

[1] Participants were informed that normally a statistical test would be required, but for simplicity, they were to make their judgments using the standard that differences were significant if greater than or equal to 5, and not significant otherwise.

format are shown in Figure 2. In the treatment and interaction sections, accuracy was highest for graphs, lower for tables, and lowest for verbal, while in the secondary section, accuracy showed the opposite trend.

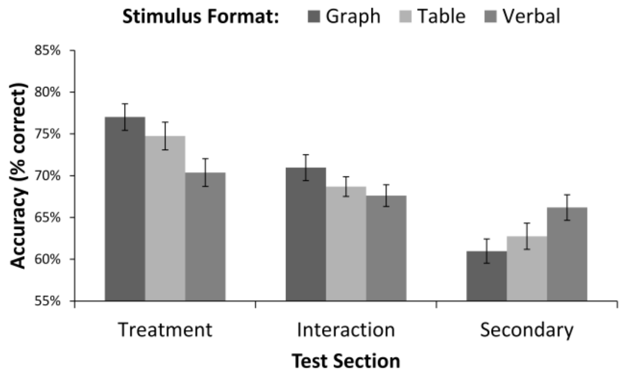


Figure 2: Accuracy by Test Section and Format. Error bars indicate standard errors.

Several of our research questions relate to graphs and tables only. Thus, the above analysis was repeated with the data from verbal stimuli excluded. The interaction of format with section was still significant, $F(2,248)=3.61$, $p=.029$. While accuracy decreased across the three sections for both graphs and tables, it decreased more for graphs (treatment: 77%, interaction: 71%, secondary: 61%) than for tables (treatment: 75%, interaction: 69%, secondary: 63%).

Response times for test trials were analyzed using the same ANOVA model structure. The results strongly resembled those for accuracy. No significant effects involving training were found, ps>.25. The main effects of test section and format were both significant, $F(2,248)=63.78$, $p\approx.000$ for section and $F(2,248)=40.69$, $p\approx.000$ for format, as was their interaction, $F(4,496)=3.04$, $p=.017$. Response times by section and format are shown in Figure 3. Responses sped up over the course of the three test sections. Responses were, overall, faster for graphs than for tables and verbal, but these differences were more pronounced in the treatment section than in the later sections.

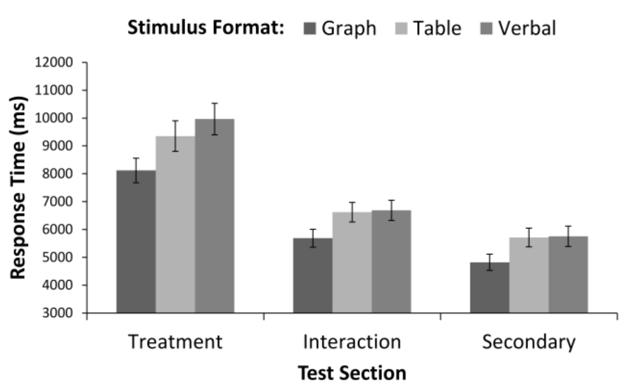


Figure 3: Response Time by Test Section and Format. Error bars indicate standard errors.

Just as for accuracy, the analysis of response time was repeated using for graph and table trials only. The main effect of format was significant, $F(1,124)=53.41$, $p\approx.000$, but the interaction of format with section was not, $F(2,248)=.913$, $p=.403$. Thus, response times were faster for graphs (6209 ms) than for tables (7230 ms) across all three sections.

Accuracy scores reflect the differing utilities of graphs and tables for task performance in different test sections, but give little insight regarding the mental processes underlying task performance. One way in which the latter might differ is the degree of influence exerted by different stimulus features. Each test stimulus was determined by presence or absence of treatment, interaction, and secondary effects, which may be viewed as three binary features. In each test section, only one feature was relevant, but the two irrelevant features may also have influenced responses. For example, in the secondary effect section, only the presence/absence of secondary effects was relevant, but a participant who had not adequately differentiated the three effects might give a positive response to a stimulus exhibiting treatment and interaction effects, even if no secondary effect was present. Thus, it could be useful to understand the influences of relevant and irrelevant features on responses for different stimulus formats and test sections.

To this end, a measure of the degree $I_{x,s}$ to which the presence of effect x influenced responses in the test section regarding effect s was calculated as follows:

$$I_{x,s} = P(R = +|E_x = +, S = s) - P(R = +|E_x = -, S = s)$$

R=+ signifies a positive response, E_x=+ and E_x=– signify, respectively, the presence and absence of effect x, and S=s signifies that the test section concerns effect s. Thus, $I_{x,s}$ represents the difference in probability of a positive response regarding effect s when effect x is present, relative to when effect x is absent. For a perfect responder, we would have $I_{x,s}$=100% when x is relevant, i.e. x=s, and $I_{x,s}$=0% when x is irrelevant, i.e. $x \neq s$. In other words, perfect responses would reflect total influence of relevant features and zero influence of irrelevant features.

Influence $I_{x,s}$ was calculated separately for each participant, stimulus format, effect x, and test section s. The pattern of results for relevant features closely resembled those for accuracy, and thus are not reported here. The results for irrelevant features are shown in Figure 4. The mean of $I_{x,s}$ in these cases was 18%, and was significantly greater than 0% for all combinations of format and test section. Thus, participants were significantly biased towards positive responses by the presence of irrelevant features.

The data for influence $I_{x,s}$ over all cases where $x \neq s$ were analyzed using the same ANOVA model structure as for accuracy and response time. No significant effects involving training condition were found, ps>.12, nor was the main effect of test section significant, $F(2,248)=0.86$, $p=.423$. However, a significant main effect of format was found, $F(2,248)=5.68$, $p=.004$, indicating that irrelevant features had the most influence for graphs (20.5%), less for tables

(18.7%), and least for verbal format (17.4%). This effect was qualified by a format × test section interaction, $F(4,496)=8.61$, $p\approx.000$. As shown in Figure 4, the influence of irrelevant features increased over test sections for graphs, stayed about the same for tables, and decreased for verbal stimuli. Separate ANOVAs conducted using the data for each format alone found a significant effect of test section on influence $I_{x,s}$ for graphs, $F(2,248)=8.41$, $p\approx.000$, but not for tables or verbal stimuli, ps>.38.

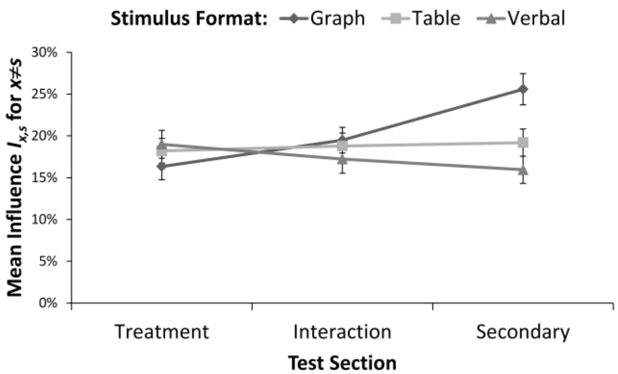


Figure 4. Influence $I_{x,s}$ for $x \neq s$, i.e. for irrelevant features. Error bars indicate standard errors.

Discussion

In the first two sections of the study, participants were trained to judge whether treatment and interaction effects were present in bivariate data, and then tested on their ability to do so when the data was presented in graphical, tabular, or verbal format. In both sections, responses were faster and more accurate for graphs than for tables. Judging the presence of either effect requires assessing complex relationships between data points, i.e. comparing averages of pairs of data points for treatment effects or differences between pairs of data points for interaction effects. The advantage shown by graphs over tables is thus consistent with the general view that complex relationships among data points are more easily assessed in graphical than in tabular format (Meyer et al., 1999; Porat et al., 2009; Schonlau & Peters, 2012; Vessey & Galletta, 1991).

Accuracy was lower in the secondary effect section than in the previous two sections. This result is not surprising, considering that participants were not told how to judge the presence of secondary effects. However, interestingly, the effect of format on accuracy was reversed in this section. What might have caused this reversal? One possible explanation, detailed below, involves transfer. Specifically, low accuracy with graphs in the secondary effect section may have reflected negative transfer from the previous sections that was absent, or reduced, in the case of tables.

To flesh out this possibility, we consider how experience of the earlier sections of the study might have affected performance in later sections. In the earlier sections, participants were trained in explicit calculation methods to judge the presence of treatment and interaction effects. With graphs, however, their judgments may have relied in part on visual patterns. For example, a sideways “v” shape in the graphs (Figure 1a) could be a useful cue for the presence of both treatment and interaction effects. Reliance on such visual patterns may have led to the creation of automatic visual routines (Ullman, 1984) that could support quick judgments regarding presence or absence of effects without, or before, performing the relevant calculations. Importantly, such routines, once acquired in the earlier sections of the study, might continue to be used in the later sections.

Thus, visual routines associating responses with visual patterns are one mechanism by which experience of the earlier sections might influence performance in the later sections. Importantly, this account predicts that such influence would be greater for graphs than for tables. Visual patterns are believed to play an important role in graph comprehension (Carpenter & Shah, 1998; Pinker, 1990), but are much less salient in the case of tables. Moreover, the above mechanism could lead to negative transfer. Because visual patterns that were relevant earlier become irrelevant, even misleading, later, continuing to rely on them could hurt performance. For example, having learned in the first two sections to give positive responses when seeing the sideways “v” shape (Figure 1a), participants might continue to do so in the secondary effect section, even though that shape actually indicates the *absence* of a secondary effect. In sum, the above account predicts greater negative transfer for graphs than for tables in the later sections of the study.

Support for this explanation comes from our analysis of influence of irrelevant features on responses. In general, such influence was greater for graphs than for tables. More important for our present purpose, such influence increased over the course of the study for graphs, exactly as would be expected if responses in later sections were influenced by visual patterns which had proven useful in earlier sections. By contrast, influence of irrelevant features did not change over the course of the study for tables, as one would expect given the lesser salience of visual patterns in tables.

An alternate explanation for the reversal, in the secondary effect section, of relative accuracies for graphs and tables involves variation in the intrinsic difficulty of recognizing different effects in different formats. Specifically, for graphs, treatment and interaction effects may have been relatively easy to detect, and secondary effects relatively difficult, while for tables, there may have been less variation in the ease of detecting the various effects. This possibility is consistent with the hypothesis, stated in the Introduction, that performance asymmetry between tasks should be greater for graphs than for tables, due to greater representational asymmetry between variables in the former case. It is also consistent with Shah and Freedman’s (2011) abovementioned finding that spontaneous interpretations of bivariate graphs tend to focus on main effects of the legend variable (in our study, the treatment factor) rather than the x-axis variable (in our study, the secondary factor).

However, two aspects of our results cannot easily be explained in terms of variation in intrinsic task difficulty. The first is the observed pattern of response times. Although accuracy in the secondary effect section was lower for graphs than for tables, reaction times showed the opposite trend, i.e. faster responses for graphs. These faster responses are consistent with reliance on automatic visual routines, as described above, but less consistent with the assumption that the task was more difficult to perform with graphs. Second, variation in intrinsic task difficulty cannot explain why influence of irrelevant features increased over the study for graphs, but not for tables. However, this effect is predicted by the first account given above.

While available evidence favors the first over the second account, further research could more definitively disambiguate between them by placing the secondary effect section at the beginning, and the treatment effect section at the end. If the first account, in terms of learned visual routines, is correct, then whichever section comes last should show negative transfer for graphs. If the second account, in terms of intrinsic task difficulty, is correct, then performance on the secondary effect section should be worse for graphs regardless of when it is encountered.

Another question investigated in our study was whether comparing graphs and tables of the same data during training, as in the Comparing Representations condition, would facilitate learning and transfer. However, this prediction was not confirmed. Accuracy showed no effect of training condition, suggesting that the Comparing Representations condition was not more effective overall. Nor did accuracy show any interaction of training condition with either format or section, suggesting that the Comparing Representations condition did not produce any particular benefits for transfer, either to a novel format, i.e. verbal, or to a novel effect type, i.e. secondary effect.

Importantly, this negative finding does not address the issue of whether the use of multiple representations during instruction can benefit learners, because multiple representations were included in all of our training conditions. However, our findings do suggest that the specific technique of comparing different representations of the same data may not produce any incremental learning benefit. This finding stands in contrast to the considerable learning benefits that can result from comparing semantically different instantiations of the same concept (Gentner et al., 2003).

In conclusion, our findings are consistent with previous research in finding an advantage for graphs over tables for tasks involving complex relationships between data points. Theories of graph comprehension suggest that salient visual patterns in graphs may underlie this advantage. However, a novel finding of the present study is that such visual patterns may not always be helpful. In particular, when performing novel tasks, graph readers may focus on visual features which were relevant to previous tasks, and have difficulty shifting perspective to focus on features which were previously irrelevant. By contrast, such shifts of perspective may be relatively easier with representational formats in which visual patterns are less salient, such as tables. These considerations suggest that graphical presentation may be preferable for performing well-practiced tasks which are known in advance, while tabular presentation may be most suitable when performing or learning to perform unfamiliar tasks.

Acknowledgements

This research was in part supported by National Science Foundation REESE grant #0910218 and US Department of Education grant # R305A1100060.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, *4*(2), 75–100.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–405.

Meyer, J., Shamo, M. K., & Gopher, D. (1999). Information Structure and the Relative Efficacy of Tables and Graphs. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *41*(4), 570–587.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial Intelligence and the Future of Testing* (pp. 73–126). Hillsdale, N.J.: Erlbaum.

Pomerantz, J., & Portillo, M. (2012). Emergent Features, Gestalts, and Feature Integration Theory. In J. M. Wolfe & L. Robertson (Eds.), *From Perception to Consciousness: Searching with Anne Treisman* (pp. 187–192). Oxford University Press.

Porat, T., Oron-Gilad, T., & Meyer, J. (2009). Task-dependent processing of tables and graphs. *Behaviour & Information Technology*, *28*(3), 293–307.

Ratwani, R. M., Trafton, J. G., & Boehm-Davis, D. A. (2008). Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology: Applied*, *14*(1), 36–49.

Schonlau, M., & Peters, E. (2012). Comprehension of Graphs and Tables Depend on the Task: Empirical Evidence from Two Web-Based Studies. *Statistics, Politics, and Policy*, *3*(2).

Shah, P., & Freedman, E. G. (2011). Bar and Line Graph Comprehension: An Interaction of Top-Down and Bottom-Up Processes. *Topics in Cognitive Science*, *3*(3), 560–578.

Ullman, S. (1984). Visual Routines. *Cognition*, *18*(1-3), 97–159.

Vessey, I., & Galletta, D. (1991). Cognitive Fit: An Empirical Study of Information Acquisition. *Information Systems Research*, *2*(1), 63–84.