Data Visualization Literacy and Visualization Biases: Cases for Merging Parallel Threads

Hamid Mansoor*

Lane Harrison†

Worcester Polytechnic Institute

ABSTRACT

People are prone to a litany of biases when viewing data visualizations. Recent visualization research has uncovered biases that manifest during visualization use, quantified their impact, and developed strategies for mitigating such biases. In a parallel thread, visualization research has begun to investigate how to measure a person's data visualization literacy, and examine the performance consequences of individual differences in these literacy measures. The aim of this position paper is to make a case for merging these threads. To bridge the gap, we highlight research in cognitive biases, that has established that there are relationships between the impact of biases and factors such as experience and cognitive ability. Drawing on prior work in visualization biases, we provide examples of how visualization literacy measures may have led to different results in these studies. As research continues to identify and quantify the biases that occur in visualizations, the impact of people's individual abilities may prove to be an important consideration for analysis and design.

Index Terms: Visualization, Bias, Literacy, Individual Differences

1 INTRODUCTION

Data visualizations represent complex information to aid people in activities including exploration, analysis, and decision making. Given that visualizations rely on visual cues and involve factors like uncertainty and risk, data visualizations are vulnerable to many perceptual and cognitive biases that humans are well known to be susceptible to. These biases may lead people to come to the wrong conclusion about data, and possibly towards commiting irrational acts. Biases then in a sense may render data useless, as the relatively objective aspects of "data" is replaced with systematic and sometimes unpredictable results from biases. Overall, this problem is particularly an issue in today's world, given the growing role of data visualizations in people's day to day lives.

Data visualizations are used in everything from political analyses to consumer websites, financial tools and more. Given the sheer number and diversity of people who use visualizations, individual differences may have a large impact on how effectively viewers read and use data visualizations. One important individual difference in this respect is data visualization literacy, *i.e.* measures of how proficient people are at reading charts and graphs. There are unique challenges in measuring data visualization literacy, in part due to factors such as the many types of visualizations available, the large number of possible tasks that can be performed on a visualization, or the actual data represented in the visualization. Even with fixed visualizations and tasks, there are challenges related to choosing an appropriate metric to represent literacy, such as a score, percentile, a grade, *etc.*. Researchers have begun to address this gap, by developing and evaluating measures of data visualization literacy [4, 16].

In parallel, there has been a growing interest and notable developments in research at the intersection of biases and data visualization. Works in this area span a range of initiatives, including indentifying biases that manifest in visualizations [8–10], quantifying the impact of biases on visualization task performance and design expectations [19], and mitigating biases as they occur [11, 20]. In particular, there are calls for increased attention on the methods and factors researchers use when evaluating biases in visualization, given these are still being uncovered and quantified (*e.g.* [8–10].

The aim of this position paper is to make a case for merging the parallel threads of data visualization literacy and visualization biases. In doing so, we highlight research in cognitive biases [5, 18], focusing on studies which have established that cognitive ability and experience can play a role in how susceptible a person is to a particular type of bias. The results, research methods, and organizational frameworks from these prior works may provide the visualization community with new means for investigating biases in data visualizations, for example by placing more emphasis on how variation in the impact of bias may be related to variations in human abilities such as their ability to inhibit biases, or by establishing that some biases are inevitable regardless of a person's individual experience and ability. Merging the data visualization literacy and visualization bias threads may also bring implications for visualization design, such as highlighting pitfalls for using more complex visualization types to mitigate biases, given that users with low visualization literacy or experience may have trouble using them.

To illustrate how visualization literacy and biases may interact, we revisit prior work on visualization biases. For example, we cover studies on the attraction bias and availability bias from Dimara *et al.* [9, 10], and discuss how data literacy measures could add dimensions and potentially impact their analyses and resulting discussions. We also cover studies that propose the use of visualizations to mitigate bias, such as Dragicevic *et al.* [11], and show how results in visualization literacy [16] may mediate the effectiveness of proposed visualizations to mitigate biases. Taken together, these examples imply that as data visualization research continues to identify and quantify the biases that occur in visualizations, the impact of people's individual abilities may prove to be an important consideration for analysis and design.

2 BACKGROUND

Prior work in visualization literacy spans different research communities. Specifically, prior work has spanned beyond the information visualization community (*e.g.* [12, 13, 16]) including communities such as intelligent tutoring systems [3] and K12 education [2, 14, 21]. The background discussed here covers work from these areas, in particular focusing on developments in visualization literacy that may relate to research targeting visualization biases.

2.1 Measuring Data Visualization Literacy & Quantifying its Impact on Performance

Recent work in data visualization literacy has focused on the accurate assessment and representation of visualization proficiency. As these

*e-mail: hmansoor@wpi.edu

†e-mail: ltharrison@wpi.edu

measures are developed, they are often accompanied by studies which illustrate the impact of high or low visualization literacy on tasks involving data visualizations.

Boy *et al.* [4] introduce a principled methodology for constructing assessment tests. Their methodology provides a blueprint for designers to come up with comprehensive, scalable, and rapidly deployable literacy assessment tests. They demonstrate the use of those rules in a user study containing four tests: two for line graphs and one each for bar charts and scatterplots. Having validated the predictive quality of this test, their work may be used by visualization researchers to add a literacy assessment component to their studies quickly and with little cost.

Lee *et al.* [16] propose the Visualization Literacy Assessment Test (VLAT), which leverages established test design methodologies alongside input from visualization experts to assign a numerical score of visualization literacy. Specifically, the VLAT uses a six-step iterative process from Psychological and Educational Management research [6] to devise the test, with a specific focus on distinguishing expert visualization users from novices. An important consideration in their design was the range of possible tasks. Given a scatterplot, for example, participants are asked questions not only about individual points, but also comparisons between points and trends. The VLAT's use of a range of visualization tasks helps ensure that it can gauge a broad spectrum of participants' abilities with visualization. In a crowdsourced study, participants took the VLAT and then attempted a few questions about an unfamiliar data visualization, a Parallel Coordinates Plot (PCP). The results of this study indicated that participants who score high on the VLAT were more likely to perform well with a visualization unfamiliar to them.

Tests that assess and quantify visualization literacy may help designers gain an idea of how their target audience's proficiency aligns with their own, which can lead to more effective visualization designs. Additionally, researchers can readily add these tests to their on visualization experiments, with relatively little cost in terms of participant effort or analysis time. Future research surrounding the assessment of visualization literacy may continue to cover more visualization tasks and contexts, given that everyday people are viewing visualizations at a greater rate than ever before.

2.2 Novices, Experts, and Visualization Use

Several studies have examined the thought process behind novices' interpretation and creation of data visualizations. Such studies aim to develop models or frameworks to capture that thought process for visualization creators to consult in their design process. While the studies reported here focus on novices and visualization use or construction, they also highlight potential biases that users face when having little experience with a visualization type or task.

Lee *et al.* [15] capture the novice thought process with the NOVIS model, which details five steps in which novices read data visualizations including: encountering the visualization, constructing a frame, exploring the visualization, questioning the frame, and floundering on the visualization. To develop NOVIS, their study included asking participants, who were college students, questions about three unfamiliar visualizations (parallel coordinates plots, chord diagrams, and treemaps), followed by asking them to verbalise their approach as they navigated through the charts. Students' comments were then used to infer and generalize five stages of how novices approach unfamiliar charts.

Beyond visualization use and interaction, research has also focused on how novices *create* visualizations. For example, a study from Grammel *et al.* [12] aimed at investigating the barriers that novices face when creating visualizations. In a user study, participants were asked to generate data visualizations through a mediator using Tableau, and to verbalise their thoughts while generating the visualizations. Novices were reportedly unable to consistently specify visualizations, and indicated that their preferences were influenced by their experience with prior data visualization types. The results of this think aloud led to the proposal of three main barriers to visualization creation: selection, visual mapping, and interpretation.

The education community has also studied how novices read charts. A study from Baker *et al.* [2] examined how K12 students interpreted and generated data visualizations. They presented middle school students with exercises to generate histograms, scatterplots, and stem-and-leaf plots, capturing their design and construction process. Students reported little experience with histograms, scatterplots, and stem-and-leaf plots, but had considerably more experience with bar charts as a result of their schools' curriculum. The study found that the generation, interpretation, and selection of the new visualizations were heavily influenced by the transfer of prior experience of bar charts. The researchers demonstrate that this bias may have been dependent upon surface similarities between bar charts and the other charts.

Kwon *et al.* [13] studied how effective different tutorial techniques are for teaching visualization skills. With the goal of having participants become proficient with Parallel Coordinates Plots, they constructed multiple tutorials. A baseline condition contained no tutorial. A static tutorial included descriptions of parallel coordinates plots along with screenshots. A video tutorial showed participants how to navigate a parallel coordinates plot. Finally, an interactive tutorial allowed users to draw parallel coordinates, enter values, and interact with the chart they make. It also gave users a list of tasks to complete, providing feedback when the users were unable to correctly finish a task. The results of a user study found that participants who saw the video and interactive tutorials fared better than the baseline and static tutorials. This study suggests that skill with a visualization can be learned, and learned relatively quickly with proper tutorial methods.

2.3 Biases and Data Visualization

People are prone to many types of biases when using data visualizations. Biases that manifest in data visualizations can impact a person's performance with visualization, and possibly lead to errors in decision making tasks related to the underlying data. This growing research area focusing on biases in visualization has investigated areas such as the mitigation of biases [8], assessing the impact and prevelance of specific biases [9], and developing frameworks and approaches to analyse biases in visualizations [20]. Results from studies on visualization bias in data visualizations can improve the ways in which visualizations are designed, benefitting the visualization community at large by enabling guidelines for less error-prone transfer of information.

3 Individual Differences and Bias: Guiding Results and Organizational Frameworks

Taken together, the threads of research in data visualization literacy and visualization biases have several parallels. Studies in data visualization literacy have uncovered biases that manifest through unfamiliarity with a visualization, for example, which may not happen when experts use the same visualizations. Beyond these threads, research in cognitive psychology has focused on the systematic study of the relationship between biases and individual differences. From the visualization perspective, Peck *et al.* discusses some possible implications of linking individual differences with factors such as experience and bias [17]. Here we highlight some of the extant research in biases and individual differences, focusing on results and organizational frameworks which may inform future studies in visualization literacy and bias.

In a series of experiments, Stanovich and West [18] studied the relationship between measures of cognitive ability and known biases. For cognitive ability, they adopted the SAT (Scholastic Aptitude Test) scores of their participants, who were primarily students. They used established bias experiments, including studies on base-rate

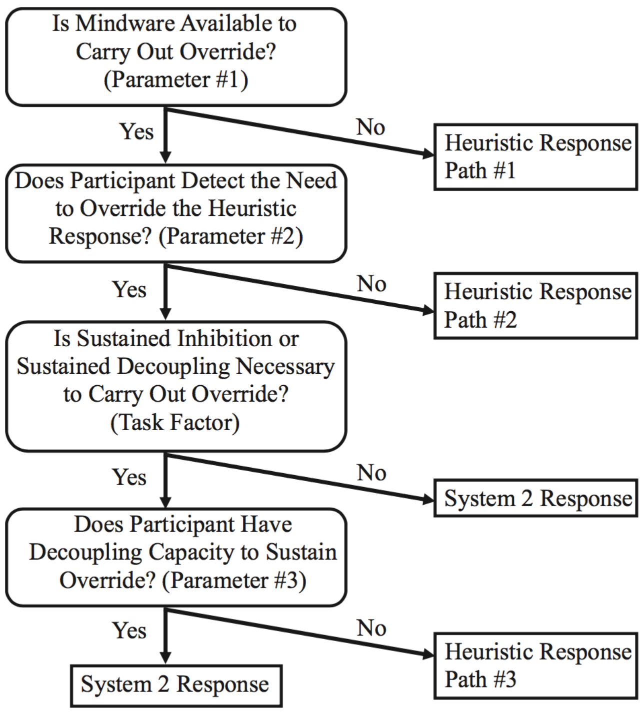

Figure 1: Stanovich and West propose an organizational framework for reasoning about when individual differences may play a role in a person's ability to mitigate biases [18]. Given known effects of experience, ability, and bias, we propose that similar measures be adopted and used in the study of biases that occur in visualization use.

neglect, anchoring effects, outcome bias, and more. The results of their experiments indicated that some of these were uncorrelated with participants' cognitive ability. Others, however, did show an effect. To reconcile this difference and provide guidance for future experiments, Stanovich and West propose a "mindware" organizational framework which illustrates the ways in which ability may or may not impact performance in bias-prone tasks. This framework is shown in Figure 1[1]. Their overall conclusion was that a person with high cognitive ability may be more able to take extra measures to prevent bias-induced errors if they are informed beforehand that the task they are about to perform involves a particular type of bias.

We know from extant research that viewing visualizations may result in many sorts of biases, but are all people equally susceptible? Stanovich and West [18] offer examples of how long term exposure and training with statistics and probabilities may equip people with mindware that allows the triggering of an altering response, also known as a bias inhibiting response. As the visualization community continues to quantify the impact and ways of measuring visualization literacy, it is possible that long term exposure to visualizations and deliberate practice with visualizations may result in people developing bias inhibiting responses for biases that occur when viewing data visualizations.

4 Linking Data Visualization Literacy to Existing Studies of Visualization and Biases

Given evidence from prior work that individual differences in visualization literacy can impact people's performance with data visu-

[1] Note: figures in this paper have been reproduced from the original publications. After review, the authors will either obtain permission for these figures, or generate alternate versions manually.

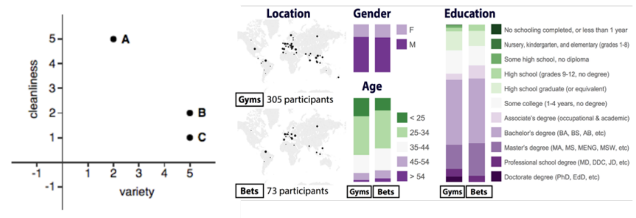

Figure 2: (Left) Dimara *et al.* establish the attraction bias in visualizations such as scatterplots, where a decoy point can systematically bias participant choice. (Right) While their study includes measures of education and other individual differences, measures of cognitive ability have been shown to play a role in biases. Newly developed visualization ability assessments [4, 16] may add informative dimensions to bias studies such as these.

alizations, that indvidual differences can play a role in the impact of biases, and an organizational framework for thinking about this interplay [18], we now consider how results and methods from data visualization literacy research could be integrated into existing studies of visualization and biases.

The Availability Bias: Dimara *et al.* [10] examines the availability bias and how it may manifest in visualization. They present a political voting decision as an example of a process that can fall prey to the availability bias. To mitigate the availability bias in this situation, they propose three ways in which data visualizations can help, focusing on how visualizations can aid recall to remove biases, and how heuristic inspired visualizations may be able to strike a balance between simplicity and accuracy to aid visualization users in avoiding biases. Following these mitigation strategies, they suggest that imperfections in visualizations can be tolerated if they increase understanding, a sentiment echoed by Correll and Gleicher [7]. What is less understood, however, is how factors like imprecise representations and complexity are related to variations in user ability. To give a concrete example, a complex visualization that mitigates the availability bias in the hands of an expert may be fine, but studies from Kwon *et al.* [13] note that novices enter many distinct stages when learning a new visualization. Thus, the populations used when testing visualization mitigation strategies may need to be taken into account, to ensure validated mitigation strategies perform as expected with potential end-users.

The Attraction Bias: In another study, Dimara *et al.* [9] examined the prevalence of the attraction bias in data visualizations. The attraction bias occurs when a viewer's choice of two options is influenced by an irrelevant third choice, *i.e.* a choice that is clearly inferior to both the original ones. To test the extent to which the attraction bias manifests in visualizations, they conducted a user study with tables (as a baseline) and scatterplots. The results of this study indicated that data visualizations are also prone to this bias, causing people to make errors in judgement.

These results may benefit by adding literacy measures as a factor. Because this study involved participants of varying backgrounds, the inclusion of literacy assessment tests such as those proposed by Boy *et al.* [4] or Lee *et al.* [16] could potentially uncover additional signal in the data. Specifically, participants in the attraction bias study gave information about their education background and confidence in their choices. Measures of visualization literacy would operate similarly to measures of demographics. For example, a numerical score of literacy could readily be factored into a correlation calculation with error. Such an analysis may identify subpopulations that are more or less susceptibile to attraction biases, or establish the robustness of the bias to people of varying backgrounds. Further, frameworks for reasoning about individual differences and biases (*e.g.* Figure 1) could be used in discussions to reason about whether mitigation strategies are possible for a given bias.

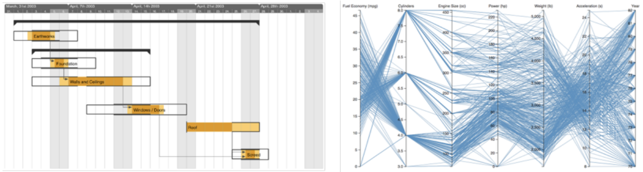

Figure 3: (Left) Dragicevic *et al.* [11] propose visualizations like PlanningLines [1] as a means for mitigating the planning fallacy. (Right) However, Lee *et al.* found that the efficacy of more complex visualizations (such as parallel coordinate plots), are significantly modulated by visualization literacy scores [16], suggesting that the use of more complex visualizations may lead to additional performance costs.

THE PLANNING FALLACY: In contrast to examining biases that manifest through visualization use, Dragicevic *et al.* [11] proposed using data visualizations to mitigate the planning fallacy, a common bias in which people make estimations of the time it would require them to finish a project. They propose four ways in which visualizations may help prevent the planning fallacy. Namely, by providing: increased awareness of obstacles, selflogging of durations and predictions, tools for supporting group predictions, and social networking tools. Their discussion moves beyond the individual and into teams, as team projects may be more susceptible to planning fallacies, given that team members are often unaware of the each other's schedules and full capabilities. As a possible mitigating visualization meeting some of these criteria, they describe Aigner *et al.*'s PlanningLines [1] (shown in Figure 3).

The interplay between complex visualizations and visualization literacy could apply in this case. Specifically, Lee *et al.* [16] showed in their study evaluating the Visualization Literacy Assessment Test (VLAT) that the scores on the VLAT were positively associated with peoples' ability to navigate unfamiliar visualizations. While the VLAT study focused on parallel coordinates plots, the proposed PlanningLines visualization uses a variety of visual encodings: glyphs, links, overlapping bars for uncertainty, *etc.*, which could imply that experience or training are necessary to achieve the goal of mitigating the planning fallacy. PlanningLines is an unfamiliar visualization and may cause issues for novices in a group to misinterpret it. Further, as the tool is proposed to be used in teams, it is unknown how individual differences such as visualization literacy will manifest when people with multiple different abilities are modifying and interpreting the same visualizations.

4.1 Reversal: Augmenting Data Visualization Literacy Research with Biases

Thus far, we have discussed how incorporating literacy as a factor in studies concerned with biases in visualizations may lead to deeper or different conclusions. However, we also note that the ongoing development of visualization literacy assessment approaches can be informed by research in visualizations and biases. Specifically, if it can be shown that certain biases that manifest in visualizations can be inhibited through experience and training, assessment questions that include such bias-prone tasks may prove useful in discriminating between novice and experienced users.

5 CONCLUSION

The aim of this position paper is to make a case that research in data visualization literacy and biases in visualization are two parallel threads that, when merged, may reveal new insights about one another. In making this case, we draw on work from cognitive psychology that has established that individual differences and biases are often (though not always) related, and cover one of the resulting organizational frameworks for thinking about the conditions under which visualization biases might be mitigated. We revisit several studies on visualizations and biases, and tie them to extant works in data visualization literacy. As these related areas continue to grow, mutual consideration may prove beneficial in furthering our understanding of visualization analysis and design.

REFERENCES

[1] W. Aigner, S. Miksch, B. Thurnher, and S. Biffl. Planninglines: novel glyphs for representing temporal uncertainties and their evaluation. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pp. 457–463. IEEE, 2005.

[2] R. S. Baker, A. T. Corbett, and K. R. Koedinger. Toward a model of learning data representations. In *Proceedings of the 23rd annual conference of the Cognitive Science Society*, pp. 45–50, 2001.

[3] R. S. Baker, A. T. Corbett, and K. R. Koedinger. Learning to distinguish between representations of data: a cognitive tutor that uses contrasting cases. In *Proceedings of the 6th international conference on Learning sciences*, pp. 58–65. International Society of the Learning Sciences, 2004.

[4] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics*, 20(12):1963–1972, 2014.

[5] J. J. Christensen-Szalanski, D. E. Beck, C. M. Christensen-Szalanski, and T. D. Koepsell. Effects of expertise and experience on risk judgments. *Journal of Applied Psychology*, 68(2):278, 1983.

[6] R. J. Cohen, M. E. Swerdlik, and S. M. Phillips. *Psychological testing and assessment: An introduction to tests and measurement*. Mayfield Publishing Co, 1996.

[7] M. Correll and M. Gleicher. Bad for data, good for the brain : Knowledge-first axioms for visualization design. In G. Ellis, ed., *DECISIVe : Workshop on Dealing with Cognitive Biases in Visualisations*, 2014.

[8] M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE transactions on visualization and computer graphics*, 20(12):2142–2151, 2014.

[9] E. Dimara, A. Bezerianos, and P. Dragicevic. The attraction effect in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):471–480, 2017.

[10] E. Dimara, P. Dragicevic, and A. Bezerianos. Accounting for availability biases in information visualization. In G. Ellis, ed., *DECISIVe : Workshop on Dealing with Cognitive Biases in Visualisations*, 2014.

[11] P. Dragicevic and Y. Jansen. Visualization-mediated alleviation of the planning fallacy. In G. Ellis, ed., *DECISIVe : Workshop on Dealing with Cognitive Biases in Visualisations*, 2014.

[12] L. Grammel, M. Tory, and M.-A. Storey. How information visualization novices construct visualizations. *IEEE transactions on visualization and computer graphics*, 16(6):943–952, 2010.

[13] B. C. Kwon and B. Lee. A comparative evaluation on online learning approaches using parallel coordinate visualization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 993–997. ACM, 2016.

[14] V. Laina and M. Wilkerson. Distributions, trends, and contradictions: A case study in sensemaking with interactive data visualizations. In *Proceedings of the 12th International Conference of the Learning Sciences. The International Society of the Learning Sciences, Singapore*, 2016.

[15] S. Lee, S.-H. Kim, Y.-H. Hung, H. Lam, Y.-a. Kang, and J. S. Yi. How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE transactions on visualization and computer graphics*, 22(1):499–508, 2016.

[16] S. Lee, S.-H. Kim, and B. C. Kwon. Vlat: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):551–560, 2017.

[17] E. M. Peck, B. F. Yuksel, L. Harrison, A. Ottley, and R. Chang. Position paper: Towards a 3-dimensional model of individual cognitive differences. *ACM BELIV*, 2012.

[18] K. E. Stanovich and R. F. West. On the relative independence of thinking biases and cognitive ability. *Journal of personality and social psychology*, 94(4):672, 2008.

[19] T. Tenbrink. Cognitive discourse analysis for cognitively supportive visualisations. In G. Ellis, ed., *DECISIVe : Workshop on Dealing with Cognitive Biases in Visualisations*, 2014.

[20] T. Verbeiren, R. Sakai, and J. Aerts. A pragmatic approach to biases in visual data analysis. In G. Ellis, ed., *DECISIVe : Workshop on Dealing with Cognitive Biases in Visualisations*, 2014.

[21] M. H. Wilkerson-Jerde and U. J. Wilensky. Patterns, probabilities, and people: making sense of quantitative change in complex systems. *Journal of the Learning Sciences*, 24(2):204–251, 2015.