Title
How do communicative goals guide which data visualizations people think are effective?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Huey, Holly
Oey, Lauren A
Lloyd, Hannah
et al.

Publication Date
2023

Peer reviewed

How do communicative goals guide which data visualizations people think are effective?

Holly Huey*[1], Lauren A. Oey*[1], Hannah S. Lloyd[1], and Judith E. Fan[1,2]

[1]University of California, San Diego, CA, United States
[2]Stanford University, Stanford, CA, United States

Abstract

Data visualizations are powerful tools for communicating quantitative information. While prior work has focused on how experts design informative graphs, little is known about the intuitions non-experts have about what makes a graph effective for communicating a specific message. In the current study, we asked participants (*N*=398) which of eight graphs would be most useful for answering a particular question, where all graphs were generated from the same dataset but varied in how the data were arranged. We tested the degree to which participants based their decisions on sensitivity to how easily other participants (*N*=542) would be able to answer that question with that graph. Our results suggest that while people were biased towards graphs that were at least minimally informative (i.e., contained the relevant variables), their decisions did not necessarily reflect sensitivity to more graded but systematic variation in actual graph comprehensibility.

Keywords: data visualization; graph production; graph comprehension; communication; pragmatics; design

Introduction

From displays of predicted weather developments, migrations of bird populations, to consumer market trends, data visualizations are a ubiquitous tool for communicating patterns in quantitative data. Their power to do so arises from their ability to distill complex information into a format that can be readily apprehended in visual form (Franconeri, Padilla, Shah, Zacks, & Hullman, 2021; Tversky, 2001; Card, 1999; Bertin, 1983; Tufte, 1983). Critically, different data visualizations—even when generated from the same underlying dataset—can be used to highlight different kinds of information depending on the communicative context. For example, a single bar plot can be used to aggregate many observations to convey the exact magnitude of their mean. But multiple bar plots across several panels might be used to convey variation in this mean across groups within the same dataset.

Our ability to judge which plot to make depending on what information is most relevant is critical for effective communication using graphs and is so important that these judgments continue to motivate the development of practical guidelines for effective visualization design (Kelleher & Wagener, 2011; Saket et al., 2018; Ajani et al., 2021). These guidelines are often informed by our empirical understanding of constraints on human perception and information processing (Kosslyn, 1989; Shah & Hoeffner, 2002; Cleveland & McGill, 1987; Franconeri et al., 2021; Rensink & Baldridge, 2010; L. M. Padilla, Creem-Regehr, Hegarty, & Stefanucci, 2018), as well as individual differences in visualization literacy (Mansoor & Harrison, 2018; Börner, Bueckle, & Ginda, 2019; Boy, Rensink, Bertini, & Fekete, 2014; Lee, Kwon, Yang, Lee, & Kim, 2019).

However, while constraints on graph *comprehension* are often the target of empirical study, graph *production* has rarely been empirically investigated in non-practitioners (Grammel, Tory, & Storey, 2010). Nonetheless, genuine visualization literacy arguably encompasses both capacities: the ability to interpret a graph and the ability to produce an interpretable graph. Graph production itself depends on two further competencies: the ability to *generate* graphs and the ability to *evaluate* the degree to which a graph is informative. While the former poses some practical barriers—e.g., because of the technical requirements of actually creating a plot of real data—if that requirement is lifted, then it becomes feasible to investigate the evaluative judgments that are integral to graph production, and thus also visualization literacy in the broader population.

Coordinated investigation of both comprehension and production has long been a cornerstone of the study of linguistic communication (Pickering & Garrod, 2013; Clark & Hecht, 1983). Over the past several years, there have been remarkable advances in our understanding of how communicative goals and context impact the production and interpretation of linguistic utterances (Degen, Hawkins, Graf, Kreiss, & Goodman, 2020; Kao, Wu, Bergen, & Goodman, 2014; Goodman & Frank, 2016; Franke & Jäger, 2016; Grice, 1975). Together, this work has provided converging evidence that a core component of natural language use is the ability to deploy mental models of other people to disambiguate meanings (i.e., during comprehension) and to generate expectations about what will be informative to other people (i.e., during production). More recently, these insights have been successfully extended to explain key aspects of how people produce informative pictorial representations in real-time visual communication tasks (Fan, Hawkins, Wu, & Goodman, 2020), suggesting that these principles may generalize beyond the domain in which they were originally developed.

The goal of the current paper is twofold: (1) to measure people's judgments about the informativeness of a graph for answering a specific question; and (2) to test the degree to which these judgments reflect sensitivity to how easily *other* people are actually be able to answer that question with that

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

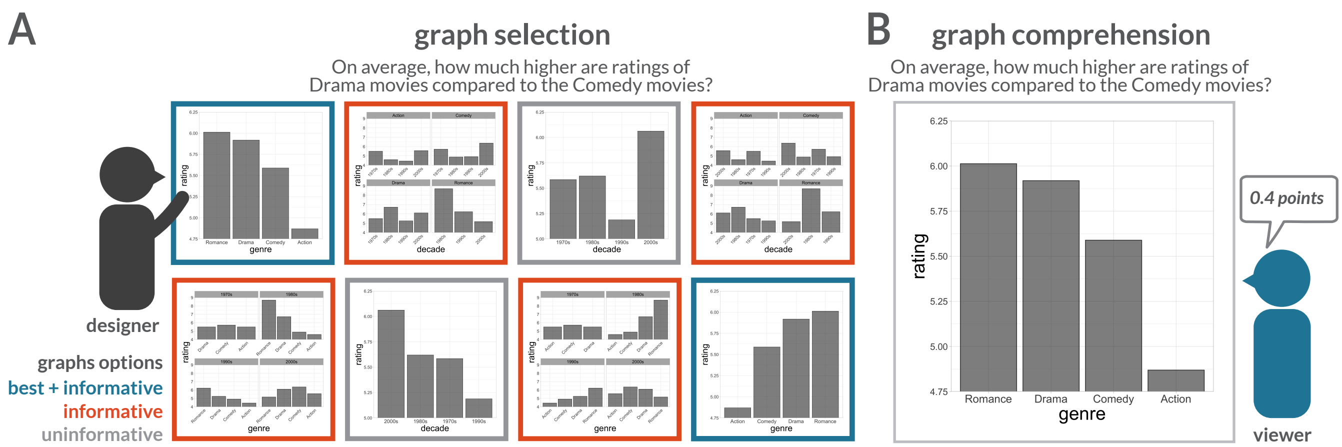

Figure 1: (A) Graph selection task. Example question prompt and corresponding graphs. Participants selected 1 of 8 graphs to help another person answer the prompted question as quickly and accurately as possible. For this example question, *best* graphs are predicted to support fast and accurate comprehension. *Informative* graphs contain the minimal information necessary to answer, whereas *uninformative* ones do not. (Graphs are color-coded for illustrative purposes of this figure, but were not color-coded in the actual task.) (B) Graph comprehension task. Viewers provided numerical answers to the same question using a corresponding graph.

graph. Towards this end, we conducted a graph selection task in which people decided which of eight graphs would be most useful for answering a particular question (e.g., "On average, how much higher are ratings of Drama movies compared to Comedy movies?"), where all graphs were generated from the same dataset but varied in how the data were arranged (Fig. 1A). Next, we conducted a graph comprehension task to obtain estimates of how well people could actually answer questions about those data visualizations. We then asked to what degree the graphs that best supported graph comprehension were also those that participants in the graph selection experiment were most likely to choose, and vice versa. We also included two heuristic baselines: first, one in which participants in the graph selection experiment chose graphs that were minimally informative (i.e., contained the relevant variables that were mentioned in the question) with equal probability; and second, one in which participants' choices reflected indifference between the graph options provided (Fig. 1B).

Hypotheses

Our aim was to evaluate how communicative goals guide how people think about what makes data visualizations informative to others. To achieve this goal, we tested three specific hypotheses:

Hypothesis 1: Audience-sensitivity If a person's judgments about data visualization design are sensitive to what naïve viewers may need to answer specific questions, we predicted that people would have strong preferences for data visualizations based on a presented question that they are tasked to help someone else answer (Fig. 2A, *left*). More concretely, we hypothesized that they would prioritize two goals: first, to identify graphs containing the minimal information necessary to answer a presented question (e.g., although a graph may be generated from an appropriate dataset, it may not contain all the information necessary to answer a specific question about it if a specific variable is not plotted); and second, among those "informative" graphs, to selectively prioritize those that would help viewers quickly and accurately interpret them.

With respect to this second hypothesis, we predicted that people would prioritize graphs that would help reduce the cognitive effort needed to extract information from them. For example, even if an informative graph may present all the information necessary to answer a specific question, it may present that information spread across multiple panels. Here, we hypothesized that this presentation style would be more cognitively taxing to a viewer, who must perform additional mental aggregation if asked to compare information across panels. Additionally, we hypothesized that graphs presenting unorganized information would be more cognitively challenging to extract highest and lowest values, relative to those that were organized by ascending numerical values. We formalized this hypothesis by developing a computational model of the graph designer that is more likely to select graphs that will lead to better comprehension by a naïve viewer.

Hypothesis 2: Minimal informativity to audiences On the other hand, if people are not sensitive to the degree of cognitive effort required by a viewer to comprehend a graph, but instead only consider whether a graph contains the minimum information needed by naïve viewers to answer specific questions about a graph (first goal of Hypothesis 1), we predicted that people would largely ignore "uninformative" graphs that omit relevant variables but would have uniform preferences among the remaining "informative" graphs (Fig. 2A, *middle* and *left*).

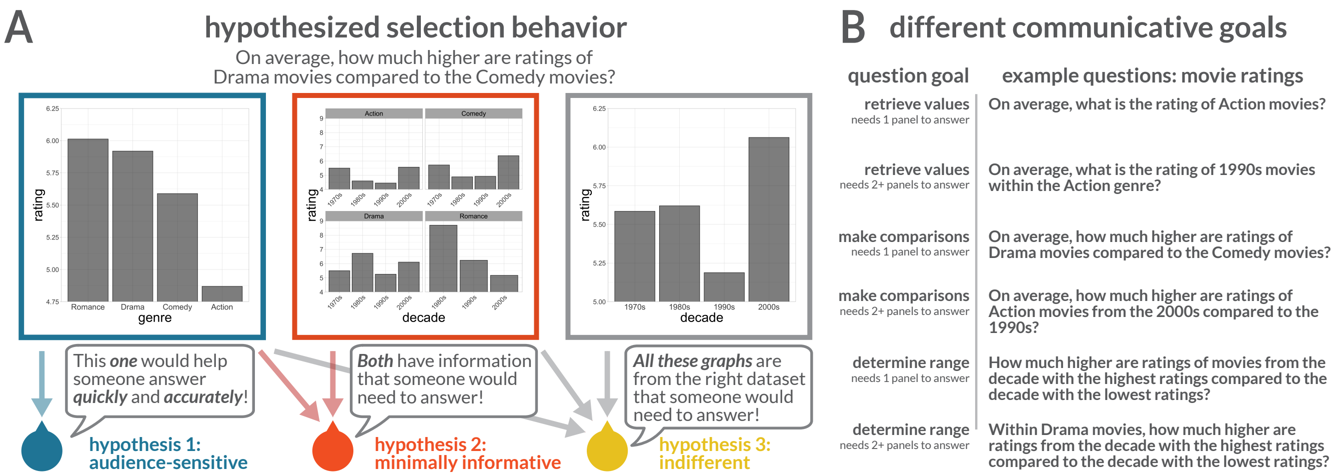

Figure 2: (A) Schematic comparison of judgments predicted by each hypothesis. (B) Example question for each question type.

Hypothesis 3: Indifference to audiences However, if people's judgments are indifferent to any communicative goals so long as a graph is generated from an appropriate dataset, we predicted that they would randomly and uniformly select from *all* presented graphs (Fig. 2A, *right*, *middle*, and *left*).

Methods

To systematically measure people's preferences for different data visualizations depending on their goal, we developed a graph selection task to measure the range of preferences that people have when trying to communicate specific information to viewers in graph form. Next, to generate behavioral predictions of our audience-sensitive model, we used a graph comprehension task to assess how well naïve viewers could quickly and accurately answer questions about those same graphs. To reduce potential unfamiliarity with different types of data visualizations, we focused on bar graphs, which are one of the most common data visualizations used in education, STEM fields, and journalistic reporting. Additionally, because prior research has suggested that focusing learning on graphing software can lead to student errors (Leonard & Patterson, 2004), we used an alternative-force choice paradigm in our graph selection task in which participants were presented with pre-generated data visualizations.

Participants 398 participants (191 male; M_{age} = 39.6 years) completed a web-based graph selection task. We excluded data sessions from 3 participants who did not complete the test trials and 7 participants who experienced technical difficulties. Another 542 participants (275 male; M_{age} = 38.4 years) completed a web-based graph comprehension task. We excluded data sessions from 6 participants who experienced technical difficulties. All participants were recruited from Prolific and provided informed consent in accordance with our institution's IRB.

Stimuli In order to generate a diverse stimuli set of bar graphs, we selected 8 popular datasets from the `MASS` package (Venables & Ripley, 2002). Each dataset contained both numerical and categorical data and were preprocessed to consist of two to four categories, so that the generated graphs would be matched in approximate visual complexity. From each dataset, we generated 8 bar graphs representing means by manipulating three commonly used parameters: (1) grouping in one or multiple separate panels (i.e., faceting), (2) x-axis variable, and (3) organization by ascending ordering of numerical x-axis variables or by alphabetical ordering. Our total test set thus consisted of 64 unique graphs. To avoid participants' requiring more advanced statistical knowledge about error as well as graph conventions in statistics, we did not include error bars. All bar graphs were gray-scaled to avoid irrelevant aesthetic preferences introduced by colors. Eight additional bar graphs were generated from the `iris` dataset for practice trials.

For each dataset, we generated 6 corresponding questions targeting different kinds of information (Fig. 2B). Half of the questions could only be answered if a graph was faceted with the correct variable on the x-axis, while the other half *could* be answered if it was faceted, but was more effective aggregated across facets, conditioned on it containing the correct variable on the x-axis. Additionally, we generated three question types asking participants to: retrieve mean values of a single category; make comparisons between the means of multiple categories; and determine the range between the highest and lowest means of categories. The syntax of each question type was standardized across datasets as much as possible.

Graph selection task Participants were presented with a random sequence of 8 trials, each corresponding to a unique dataset. On each trial, they read a description of a dataset and then were presented with the corresponding question about the dataset and a 4 × 2 gallery of 8 graphs (Fig. 1A). To ensure that participants viewed each graph, they were instructed to use their cursor to hover over each graph, which would then acquire a green border to help participants track which graphs they had "viewed". After viewing each graph, they were instructed to click the one that would best help someone else answer the question as quickly and accurately as possible. Graphs were presented in random order in the gallery, and participants could not select a graph until they had viewed all 8 graphs. The order of presented datasets, as well as the question type corresponding to each dataset, was randomized across participants. Participants also completed one practice trial to ensure that they were familiar with the web interface.

Graph comprehension task In each trial, participants were presented with a dataset description and corresponding question and provided a numeric answer using a presented graph (Fig. 1B). Participants were instructed to answer the question as quickly and accurately as they could, even if they had to guess. In addition to completing one practice trial prior to test trials, participants completed a random sequence of 8 test trials each corresponding to a unique dataset and were not told which graphs were informative or uninformative for a question prompt.

Results

Comprehension performance varies across data visualizations To validate our choice of graph stimuli, we first established that people produced more accurate question responses to informative graphs than uninformative ones. To test this, we fit a mixed effect linear regression model to predict viewer absolute error with random effects for participant and dataset. We found that viewers of the graph comprehension task produced more error in responses when presented with uninformative graphs compared to informative ones ($t = 10.31$, $p < 0.001$), confirming that informative graphs were more effective with helping viewers than uninformative ones. Furthermore, we applied a likelihood ratio test to a nested model comparison and found that the graph itself explained additional variation in responses ($\chi^2(7) = 37.96$, $p < 0.001$). This analysis provides additional evidence that graphs have gradient and variable utility for answering the same question, beyond just whether a graph was informative or not. These results provide validation that our stimuli set was diverse enough to capture response variation.

Communicative goals impact graph selection behavior Our main goal was to evaluate how communicative goals guide preferences that people have about data visualizations that convey different information. To accomplish this, our first step was to examine whether people select graphs in a non-uniform manner. Using a chi-square goodness-of-fit test, we found that graph selections were non-uniform to different data visualizations ($\chi^2(7) = 527.13$, $p < 0.001$) and were dependent on the presented question ($\chi^2(35) = 1590$, $p < 0.001$; Fig. 3A). These results suggest that participants were sensitive to how graphs vary in informativity for different question prompts.

To further explore these results, we then evaluated participants' selection behavior against the uniform selection distribution predicted by the indifferent hypothesis, which proposes that each graph has the same probability (12.5%) of being chosen. To quantify just how different people's strategies are from these proposed hypotheses, we applied a Jensen-Shannon divergence (JSD) metric. Here, if two distributions perfectly aligned, they would have a JSD of 0. We found that participants' graph selection distribution was significantly different from the indifferent hypothesis (JSD = 0.51; bootstrapped 95% CI = [0.48, 0.58]; Fig. 3B), providing evidence that people use a richer strategy when deciding how to communicate with graphs.

People prefer to select informative graphs rather than uninformative ones We next assessed how strongly participants might have been guided by communicative goals to help viewers accurately answer questions. We fit a logistic regression predicting the graph type (i.e., informative vs. uninformative) selected with random effects for participant and dataset. Consistent with our minimally informative hypothesis, we found that participants systematically chose informative graphs that contained at least the minimal amount of information needed to answer a corresponding question prompt, relative to uninformative ones ($z = 15.4$, $p < 0.001$).

To further explore the selection behavior predicted by our minimally informative hypothesis, we formalized a coarse-grain strategy of choosing minimally informative graphs as a softmax decision rule that prioritized informative graphs ($U = 1$) and discarded uninformative ones ($U = 0$). The softmax temperature was treated as a free parameter for each question. Although this model also fell short of perfectly matching the participants' selection behavior (JSD = 0.12; bootstrapped 95% CI = [0.11, 0.17]; Fig. 3B), it was significantly better matched to the selection behavior demonstrated by participants than the selection distribution predicted by the indifferent hypothesis ($\hat{\beta} = -0.06$, $t = -9.57$, $p < 0.001$). These results suggest that participants are motivated to prioritize informative graphs over uninformative ones in order to help viewers accurately answer questions.

Selection behavior explained equally well by heuristic strategy and audience sensitivity in these data Our analyses so far reveal that participants prioritize graphs that are informative enough to help viewers answer questions about them. Our main goal, however, was to investigate whether people's selection behavior can be explained by a sensitivity in preferences for graph that would support efficient graph comprehension by viewers. Concretely, we hypothesized that if participants were motivated help reduce the cognitive effort needed by viewers to answer questions by choosing an effective data visualization, we predicted that their selection behavior would more closely match the rich gradient variation in graph comprehension demonstrated by our validation results.

To compute predictions for an audience-sensitive model, we calculated the error of viewers' numerical responses from our graph comprehension task, relative to the ground truth

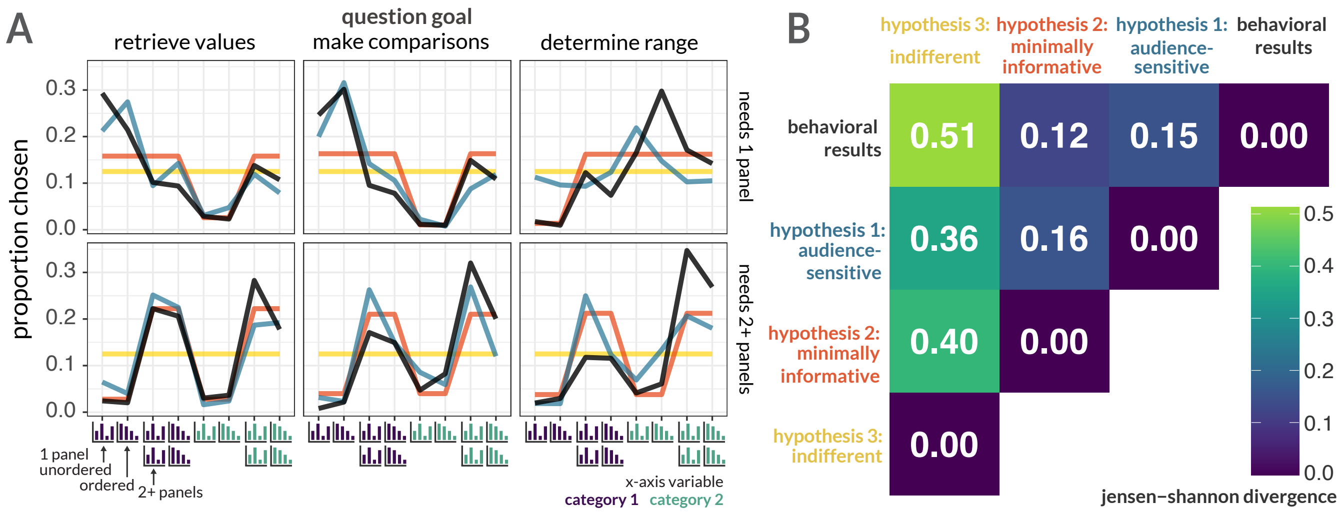

Figure 3: (A) Behavioral results (black line) compared against model predictions (colored lines), across question type and presented graph parameters. Greater overlap with the black line indicates how well the model fits the results. The behavioral results are equally well-explained by the minimally informative and audience-sensitive hypotheses. (B) Pairwise comparisons of the hypotheses, using Jensen-Shannon Divergence. Smaller numbers indicate greater similarity (identical match: 0).

answer for each dataset and question type. Aggregating over all viewers who received the same graph-question pairing, we then computed the root-mean-square-error (RMSE) wherein larger error associated with a graph demonstrates that it did not effectively help answer a specific corresponding question. RMSEs were re-scaled between 0 and 1 to normalize across the different data sets. These re-scaled RMSEs were then averaged across datasets and input into a softmax decision rule as utility values. Negative softmax temperature free parameters indicated that participants selected graphs inversely related to the amount of error that was elicited by viewers.

Next, to evaluate how well audience-sensitive selection behavior might predict viewers' ability to efficiently use a graph to answer a corresponding question, we compared this error gradient in viewer responses to the participants' selection behavior in which the JSD was found to be 0.15 (bootstrapped 95% CI = [0.14, 0.20]; Fig. 3B). Critically, while our audience-sensitive model more closely matched participants' selection behavior compared to an indifferent model ($\hat{\beta} = -0.06$, $t = -5.80$, $p < 0.001$), we did not find significant improvement in this ability to predict selection behavior compared to a minimally informative model ($\hat{\beta} = 0.005$, $t = 0.51$, $p = 0.62$). In sum, these results suggest that people are sensitive to different degrees of informativity when choosing between informative and uninformative graphs, but are not necessarily sensitive to more subtle difference between how different graphs plotting the same data could better support faster and accurate comprehension by viewers.

Discussion

Data visualization, among other tools for making sense of large volumes of data, have become increasingly important in recent decades (Holst, 2021). Here we investigated the intuitions that ordinary people have about what makes data visualizations informative for answering specific questions. Concretely, we evaluated the extent to which people may be sensitive to how communicative goals to convey different kinds of information should shift data visualization design. We hypothesized that non-experts' intuitions about effective data visualization design could be: (1) audience-sensitive to what viewers may need to accurately and efficiently interpret graphs; (2) minimally informative to audiences; or (3) indifferent to any graph, so long as it is generated from the dataset that a viewer is asked about.

To test our hypotheses, we evaluated non-experts' intuitions about data visualization efficacy using two novel tasks: First, we used a forced-choice graph selection task to remove skill-based barriers associated with graph construction (e.g., manipulating data using programming languages) and to measure participants' preferences about graphs intended to communicate different kinds of information. Next, to model audience sensitivity, we developed a graph comprehension task to evaluate a separate group of naïve viewers' ability to accurately answer questions about those same graphs. We found that people prioritized graphs containing the minimal information needed to answer prompted questions, but were not necessarily sensitive to more subtle differences between how different graphs plotting the same data could better support fast and accurate comprehension by others. Overall, our findings contribute quantitative evidence that even non-experts' intuitions about data visualization design are guided by goals to generate informative messages for others, despite lacking design expertise typically investigated by prior research. By leveraging viewers' downstream interpretations of the same graphs, our results additionally provide critical insights about how their design preferences have a direct downstream impact on viewers' ability to accurately extract information from graphs.

A key contribution of this work is that we establish the feasibility of systematically investigating non-experts' intuitions about data visualization design. How might their intuitions differ from those of experts? While our findings demonstrate that our participants were systematically biased to select informative graphs over uninformative ones, their selection behavior is consistent with a relatively coarse understanding of what makes a data visualization easy for someone else to understand. Whereas people are exposed to other forms of pragmatic communication like gestures and drawings and become adult-like experts in processing them from a young age (Goldin-Meadow, 2009; Huey & Long, 2022), experience with graphs typically occurs later in development and is taught in formal educational settings. Thus, a coarse understanding may develop into more fine-grained tuning as people gain more domain-specific experience with data visualizations or more broadly, mathematics. This prediction resonates with previous studies suggesting that graph reading performance can be predicted by learners' basic numerical abilities (Ludewig, Lambert, Dackermann, Scheiter, & Möller, 2020; M. J. Padilla et al., 1986; Berg & Smith, 1994), spatial reasoning about mental number lines (Booth & Siegler, 2008), and comprehension of non-symbolic and symbolic number magnitudes (Dehaene, Piazza, Pinel, & Cohen, 2003) and arithmetical processes (Gillan, 2009).

As an initial step to explore the relationship between expertise and data visualization design, we conducted exploratory analyses using post-test surveys in which participants' self-reported how frequently they make and interpret graphs on a weekly basis. We found that self-reports did not significantly predict participants' performance in our graph selection task (make: $p = 0.349$; interpret: $p = 0.328$) or graph comprehension task (make: $p = 0.383$; interpret: $p = 0.8002$). While our self-report results lacked variation across experience levels, real-world data or controlled interventions may be better suited to causally deduce a relation between domain-specific expertise and how people think about visual communication and statistical patterns. For example, as people gain more experience with data visualizations, they may gain greater visual acuity with discerning such statistical patterns (Ratwani & Gregory Trafton, 2008; Ali & Peebles, 2013) that go beyond ingrained Gestalt Principles.

One critical question raised by the current studies concerns how specific beliefs about data visualization efficacy may constrain people's design goals. In our graph selection task, we operationalized efficacy as accuracy under time pressures and instructed participants to choose graphs that would help other people answer prompted questions as quickly and accurately as possible. We hypothesized that because graphs have a number of unique communicative characteristics that explicitly support large-scale quantitative information compression (e.g., bar graphs represent means), participants would select graphs that could most easily and directly help others answer a prompted question. However, it is possible that people not only prioritize speed-accuracy tradeoffs when making judgments about data visualization efficacy concerning a single question but also may prioritize data visualizations that are able to answer *multiple* questions. Indeed, while data visualizations in journalistic reports may often contain less variables in order to be more comprehensible to a general audience, it may be the case that people associate more complex data visualizations as more "accurate", effective, and even scientific legitimate if more information is possible to be extracted from it. If this bias underlies people's beliefs about data visualization efficacy, this could explain why participants of the current study did not systematically select graphs that could support faster and more accurate comprehension by others. To disentangle these potential beliefs, additional work in our lab is exploring whether people are biased to select graphs with more plotted variables than the minimal variables needed to answer prompted questions.

In conclusion, our paper contributes new insights about how people transform their knowledge about the world into data visualizations that others can learn from. Indeed, research of this nature investigating graphs and their presentation of statistical patterns is critical to deepening foundational understanding of communication modalities that involve symbolic reasoning, but also to the scientific community that utilizes data visualizations as a primary tool to share findings with other scientists and with the general public. Ultimately, data visualization studies guided by cognitive theories of communication may help advance the development of novel data visualization tools, as well as identify potential opportunities for graph literacy interventions in STEM education and design.

Acknowledgments

Many thanks to the members of the Cognitive Tools Lab for helpful discussion. This work was supported by an NSF CAREER Award #2047191 to J.E.F and an NSF Graduate Research Fellowship under Grant No. DGE-1650112 to L.A.O. J.E.F is additionally supported by an ONR Science of Autonomy award and a Stanford Hoffman-Yee grant.

All code and materials available at `https://github.com/cogtoolslab/davinci_public2023`

References

Ajani, K., Lee, E., Xiong, C., Knaflic, C. N., Kemper, W., & Franconeri, S. (2021). Declutter and focus: Empirically evaluating design guidelines for effective data communication. *IEEE Transactions on Visualization and Computer Graphics*, *28*(10), 3351–3364.

Ali, N., & Peebles, D. (2013). The effect of gestalt laws of perceptual organization on the comprehension of three-variable bar and line graphs. *Human factors*, *55*(1), 183–203.

Berg, C. A., & Smith, P. (1994). Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. *Science Education*, *78*(6), 527–554.

Bertin, J. (1983). *Semiology of graphics*. University of Wisconsin Press.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child development*, *79*(4), 1016–1031.

Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, *116*(6), 1857–1864.

Boy, J., Rensink, R. A., Bertini, E., & Fekete, J.-D. (2014). A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1963–1972.

Card, M. (1999). *Readings in information visualization: Using vision to think*. Morgan Kaufmann.

Clark, E. V., & Hecht, B. F. (1983). Comprehension, production, and language acquisition. *Annual Review of Psychology*, *34*(1), 325–349.

Cleveland, W. S., & McGill, R. (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society: Series A (General)*, *150*(3), 192–210.

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A bayesian approach to "overinformative" referring expressions. *Psychological Review*, *127*(4), 591.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive neuropsychology*, *20*(3-6), 487–506.

Fan, J. E., Hawkins, R. D., Wu, M., & Goodman, N. D. (2020). Pragmatic inference and visual abstraction enable contextual flexibility during visual communication. *Computational Brain & Behavior*, *3*, 86–101.

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. *Psychological Science in the Public Interest*, *22*(3), 110–161.

Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, *35*(1), 3–44.

Gillan, D. J. (2009). A componential model of human interaction with graphs: Vii. a review of the mixed arithmetic-perceptual model. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 53, pp. 829–833).

Goldin-Meadow, S. (2009). How gesture promotes learning throughout childhood. *Child Development Perspectives*, *3*(2), 106–111.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818-829.

Grammel, L., Tory, M., & Storey, M.-A. (2010). How information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 943-952. doi: 10.1109/TVCG.2010.164

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics Vol. 3: Speech Acts* (pp. 64–75). New York: Academic Press.

Holst, A. (2021). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025. *Statista, June*.

Huey, H., & Long, B. (2022). Developmental changes in the semantic part structure of drawn objects. In *Proceedings of the 44th annual meeting of the cognitive science society.*

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*(33), 12002–12007.

Kelleher, C., & Wagener, T. (2011). Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software*, *26*(6), 822–827.

Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, *3*(3), 185–225.

Lee, S., Kwon, B. C., Yang, J., Lee, B. C., & Kim, S.-H. (2019). The correlation between users' cognitive characteristics and visualization literacy. *Applied Sciences*, *9*(3), 488.

Leonard, J. G., & Patterson, T. F. (2004). Simple computer graphing assignment becomes a lesson in critical thinking. *NACTA Journal*, 17–21.

Ludewig, U., Lambert, K., Dackermann, T., Scheiter, K., & Möller, K. (2020). Influences of basic numerical abilities on graph reading performance. *Psychological Research*, *84*, 1198–1210.

Mansoor, H., & Harrison, L. (2018). Data visualization literacy and visualization biases: Cases for merging parallel threads. *Cognitive Biases in Visualizations*, 87–96.

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, *3*(1), 1–25.

Padilla, M. J., et al. (1986). An examination of the line graphing ability of students in grades seven through twelve. *School Science and Mathematics*, *86*(1), 20–26.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347.

Ratwani, R. M., & Gregory Trafton, J. (2008). Shedding light on the graph schema: Perceptual features versus invariant structure. *Psychonomic Bulletin & Review*, *15*(4), 757–762.

Rensink, R. A., & Baldridge, G. (2010). The perception of correlation in scatterplots. In *Computer graphics forum* (Vol. 29, pp. 1203–1210).

Saket, B., Moritz, D., Lin, H., Dibia, V., Demiralp, C., & Heer, J. (2018). Beyond heuristics: Learning visualization design. *arXiv preprint arXiv:1807.06641*.

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*, 47–69.

Tufte, E. R. (1983). The visual display of quantitative information.

Tversky, B. (2001). Spatial schemas in depictions. In *Spatial schemas and abstract thought* (Vol. 79, p. 111).

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer. Retrieved from https://www.stats.ox.ac.uk/pub/MASS4/ (ISBN 0-387-95457-0)