LIMNOLOGY AND OCEANOGRAPHY
Letters

ASLO
Open Access

Limnology and Oceanography Letters 8, 2023, 213–219

doi: 10.1002/lol2.10288

ESSAY

Show and tell: approaches for effective figures

Stephen R. Midway,[1]* *Jennifer R. Brum*,[1] *Matthew Robertson*[2]
[1]Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana; [2]Centre for Fisheries Ecosystems Research, Fisheries and Marine Institute of Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada

Scientific Significance Statement

Humans are visual beings, and everything we look at initiates a cascade of cognitive reactions that hopefully ends in information transfer and understanding. Despite relatively well-understood cognitive pathways for visual messaging, scientific figures, charts, and other visuals are often developed based on familiar conventions that are disconnected from visual best practices. Fortunately, nearly all figures can be improved with a basic understanding of the characteristics of an effective figure along with the tools needed to create effective figures.

"*Trouble with you is the trouble with me,*
got two good eyes but we still don't see."

–Robert Hunter

What makes an effective data visualization?

We have all experienced the magnetic pull of colorful, clear, and informative figures that tell a scientific story. But we have also looked at scientific figures that resulted in confusion or even elicited an almost visceral negative response. Why does such a disparity of experiences exist across data visuals? The answer lies in the ways in which we process visual information and whether the figure creator has used techniques that are compatible with this process. Much of our figure-making experience develops from what we like or dislike about figures we encounter, or we come across one of the many (helpful?) articles or graphics illustrating the 'dos and don'ts' of figure-making (Kelleher and Wagener 2011; Rougier et al. 2014; Greenacre 2016; Midway 2020). Yet rarely do we think about why we are drawn to certain figures or what hidden figure-making rules may have enabled its creation. Our responses to certain figures and graphical techniques are generated by known visual cognitive processes, and such processes have been studied for decades both in the context of scientific and non-scientific communication. This article provides suggestions for making effective scientific visuals in the context of understanding *why* these suggestions elicit a more positive response from the viewer and thus increase the author's ability to communicate the scientific message.

How we understand what we see

We all use two cognitive processes to interpret visual information (Kahneman 2011; Padilla et al. 2018). The first is the fast, automatic, and impulsive visual system often referred to as System 1 decision making (or Type 1 processing; Evans and Stanovich 2013). System 1 is the reason why you can look at Fig. 1a and immediately interpret the relationship between the x and y variables. The second is System 2 decision making (or Type 2 processing)

*Correspondence: smidway@lsu.edu

Associate editor: James E Cloern

Author Contribution Statement: SRM, JRB, and MR conceived of the ideas and drafted and revised the essay.

Data Availability Statement: Data sharing is not applicable to this article as no new data were created or analyzed in this study. R scripts used to produce the figures can be found on Github: https://github.com/MatthewRobertson2452/Show-and-Tell

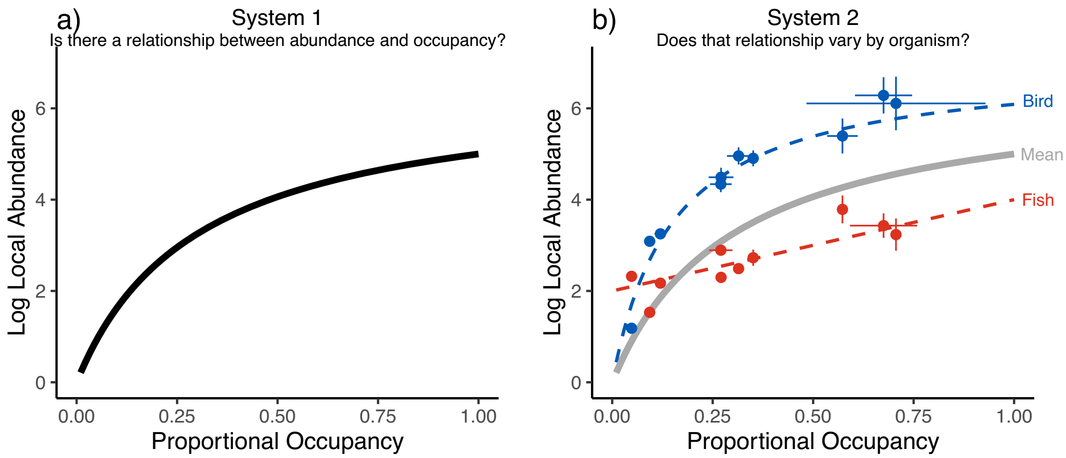

Fig. 1. Examples of figures that primarily engage System 1 decision making (**a**) and that require additional System 2 decision making (**b**). The System 1 figure represents only the idealized relationship between abundance and occupancy, requiring less interpretation from the viewer. The System 2 figure includes data and best fit lines for birds and fish that have relationships that differ from the idealized relationship and from one another, requiring more interpretation from the viewer to compare the data points and best fit lines. (All code for producing figures can be found at https://github.com/MatthewRobertson2452/Show-and-Tell)

that requires logical, conscious interpretation. When examining Fig. 1b you are required to compare the individual relationships for birds and fish to the mean of the relationships, in order to come to a conclusion about the overall data.

Anything that forces a viewer to engage in unnecessary conscious thought will reduce the effectiveness of a figure. Systems 1 and 2 operate simultaneously, and both need to be anticipated, yet the processes that initially engage a viewer are derived from System 1. For example, in Fig. 1a all extraneous information is absent, thus reducing the *foreground effect* in which the viewer may focus on only some visual information and come to an inaccurate conclusion (Stone et al. 1997). Furthermore, there is no clutter, such as a different numbers of decimal places on the *x*-axis or superficial colors. Visuals that minimize conscious thought, and therefore enhance System 1 decision making, are the reason for many of the "dos and don'ts" seen in publications relating to making scientific figures.

But we cannot always tell our story with one regression line, and as such we must consider System 2 as our visuals take on more information. Yet we still want to reduce the amount of conscious thought required for System 2 thinking. One important tactic used to achieve this is to present the data in a manner that is congruent with the viewer's expectation. First described using cognitive fit theory (Vessey 1991; subsequent review Vessey 2006), the idea of viewer expectations is congruent with System 2 decision-making in which the required amount of working memory is increased if the information is not presented in the manner expected by the viewer (Padilla et al. 2018). For example, Fig. 1b requires less conscious thought than if the same information were presented as a table because we expect information to be presented as figures when we need to interpret relationships between variables. Conversely, we expect data to be presented in a table if we need to extract individual or specific numbers. To continue our examination of Fig. 1b, there are again no distracting errors that require unnecessary cognitive processing—the focus of the viewer is on interpreting the difference between the three lines presented. You are able interpret the figure without too much thinking because (1) the bird and fish data are labeled with different words, use different line formats, and are presented using colors that differentiate them from each other (and also from the mean), and (2) you can interpret the range of data compared to the average trend lines for bird and fish because both the individual data points and average lines are presented. The ease of figure interpretation is inversely proportional to the amount of conscious thought required to interpret it and working with Systems 1 and 2 will help reduce your viewer's cognitive load.

Looking beyond the data

In addition to the dual-process systems, it is important to understand the three levels of graphical cognition that you should expect viewers to use (Friel et al. 2001). First, a viewer will attempt to *read the data*, or otherwise orient themselves to what a bar height or point size or axis means. Second, a viewer will attempt to compare the data or *read between the data*, meaning they will look for comparisons, compositions, or other relations among the parts of the figure. Finally, the highest level of graphical cognition is *reading beyond the data*, where the viewer will make inferences or predictions beyond the data presented (reviewed in Galesic and Garcia-Retamero 2011; Okan et al. 2012). The process of graphical cognition takes place within both Systems 1 and 2, albeit much faster and more intuitively in System 1. Although these three levels of graphical cognition may seem simple, they remain a good reminder and way to audit the figures you make—is a viewer able to comprehend your figure on all three levels with minimal cognitive effort? Effective visuals are those that navigate the challenges of all three cognitive levels by reducing the amount of cognitive energy it takes to complete this process and arrive at an intended message.

Recognizing if a viewer can comprehend a figure on all three levels of graphical cognition requires understanding how each level is affected by figure design. The first level of graphical cognition, reading the data, is affected by the design of axes, legends, captions, and the choice of figure used to represent the data (Fig. 2a; Boote 2014). If a viewer can't quickly and easily identify what the different figure components represent, comprehension and recall of the intended message will be hampered (Haroz and Whitney 2012; Borkin et al. 2015). The second level of graphical cognition, reading between the data, involves the recognition of relationships and interpolations between data (Boote 2014). This level of cognition is affected by the viewer's ability to interpret patterns. Pattern interpretation is made simpler by using commonly encountered figure types, making relationships easily identifiable (*see* suggestions below), and by linking the visualization to scientific concepts (Fig. 2b; Wang et al. 2012; Boote and Boote 2017). Finally, the third level of cognition, reading beyond the data, links the visualization to a hypothesis enabling interpretation of the current data and extrapolation of the message to other data (Boote 2014). The final level is mainly affected by emphasis on the message that you want the viewer to take away. In Fig. 2c, we used color and a horizontal and vertical line to place emphasis on the inflection point for population growth and on the data points that are above that level to help readers identify when population growth will decline. Emphasis created by color, shapes, and other aesthetics often improve the viewer's ability to extrapolate the figure's message and arrive at higher meaning.

Figuring out the parts

Once you have identified a simple and clear message, you need to consider the aesthetic features that will create the most direct path from your message to your viewer. This process can be challenging because (1) the best elements vary based on the data and message, and (2) there may be multiple effective elements to share your message (Hegarty 2011). Typically, the elements most manipulated are color, orientation, shape, size, length, width, and spatial position, although others exist. Figure elements can be used for different purposes within a figure; for example, some elements, like color or shape, are better for showing qualitative information, while other elements, like lengths and widths, convey quantitative information. Generally, the elements are used in combination to display different types of data and information, and their interactions should be considered. For example, a limited number of elements can be varied to emphasize the take-away message and improve recall by providing visual contrast between data sets (Rolandi et al. 2011; Borkin et al. 2015).

In general, when attempting to minimize cognitive energy with figure elements there tends to be two main guidelines: (1) simpler is better and (2) common figure types can be understood faster and more correctly than uncommon forms. "Simpler is better" can refer to reducing the number of elements used to the minimum required for the reader to understand the figure. This is because each element used in the figure (i.e., color, shape, line thickness, line type) will need to be recognized and understood by a viewer, therefore taking more cognitive energy to understand your message. However, simplifications can often go beyond removing extraneous

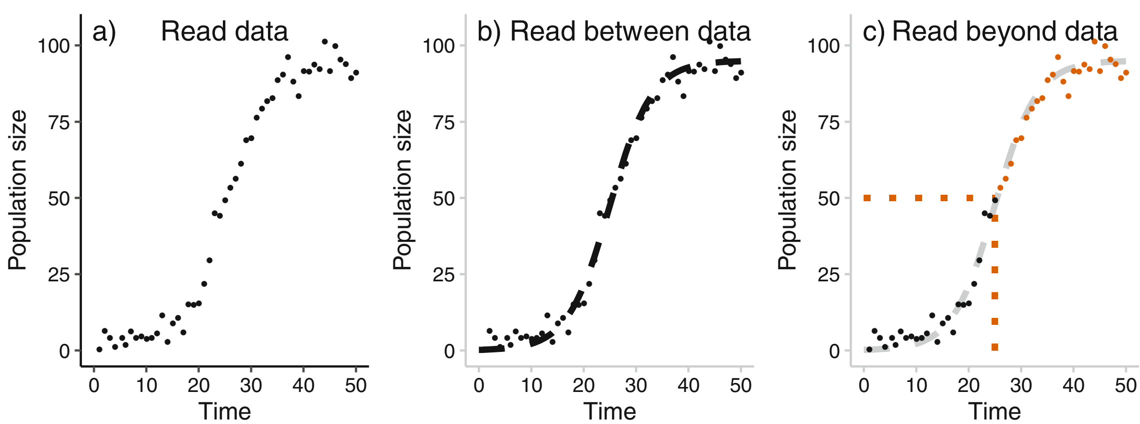

Fig. 2. Using the concept of logistic growth, a viewer will first read the data (**a**), read between the data (**b**; what pattern, relationship, or comparison is shown or suggested), and finally they will read beyond the data (**c**; what interpretation can be made). (All code for producing figures can be found at https://github.com/MatthewRobertson2452/Show-and-Tell)

Table 1. Technical recommendations for figure construction (*see* also Kelleher and Wagener 2011; Rougier et al. 2014; Greenacre 2016; Midway 2020).

Recommendation
1. All figure elements are appropriately sized and readable—fonts, labels, lines, and so on
2. Don't waste space—every element of a figure should contribute meaning or information
3. Show/include data whenever possible
4. Don't overlap points—scatter/jitter or adjust density/transparency
5. Use a legend when a plot is busy, but consider labeling elements directly if possible
6. Use titles or labels for multipanel figures
7. Consider color accessibility (e.g., colorblind-friendly; gray-scale friendly; contrasted elements; including alt text)
8. Use abbreviations sparingly and define all abbreviations
9. Use a consistent color scheme throughout the paper
10. Use small multiples for figures with the same axes; use panels if the figures are complementary

figure elements. For example, using bars to represent means of the data connects the means to the *x*-axis and causes the reader to visually search the figure space, as opposed to dots or points, which focus attention on the point of interest (i.e., the mean), thereby aiding figure interpretation (Okan et al. 2018). In addition to simplifying figures, using figure designs readers are more familiar with (such as boxplots) can reduce the amount of cognitive energy required to understand a given figure when compared to potentially more informative designs (e.g., violin plots, Anderson et al. 2011; Padilla et al. 2018). Thus, the figure design used will not only be dependent on the type of data being presented, but also on the ease in which the reader can interpret the information. A good strategy for accomplishing this is to have colleagues review the constructed figures to ensure they are clear and understandable to others, not just the author (Midway 2020). Although a full treatment of technical guidance is beyond the intention of this article, Table 1 presents some simple recommendations and sources more in-depth references for technical guidance.

Aesthetics: Keeping your figures honest

Distinguishing between cognitive and perceptual aesthetics is important because what looks better may not always improve interpretation. In his seminal work, Tufte (1983) introduced the concept of the "lie factor" to describe how well data is represented by the visualization in the figure. One of the most common ways in which interpretation of figures can be biased involves selecting an inappropriate figure type for displaying a particular type of data. For example, people tend to like presenting data in bar plots more than boxplots even though viewers are more likely to interpret bar plots incorrectly (Okan et al. 2018; *see* Box 1 for example). Another common mistake involves inappropriate axis visualization (e.g., whether or not to truncate the *y*-axis such that it does not start at zero), which have been some of the most popular sources of criticism for figure misinterpretation (Correll et al. 2020). In addition, the choice of figure elements for data representation can bias figure interpretation. For instance, modifying shape size has been shown to be more effective than modifying color when attempting to visualize differences in magnitude (Mairena et al. 2020). Finally, some elements can affect the perception of others when used in combination. For example, shape can affect a viewer's perception of size and color, and colors are more discernible in filled shapes (Smart and Szafir 2019). In summary, although there are many potential cognitive biases that may influence figure interpretation (*see* Dimara et al. 2018 for a review), the most common biases can be alleviated by choosing figure designs that match

Box 1. Building an example.
Many of the recommendations that we have provided emphasize making simple figures for improved message clarity and comprehension. However, data transparency is necessary for effective, reproducible science (Weissgerber et al. 2019). Transparency generally involves visualizing the raw data or the distribution of the data. Consider the common bar plot (Fig. 3a) and how it might be improved. Data can be added to plots in a visually simple way by layering points over previous figure elements (Fig. 3c) and by "jittering" those points along an axis to more easily discern between individual points (Fig. 3d). (Jittering is simply adding a small amount of random variation to overlapping points, in order to visually separate them.) Furthermore, separation of figures into small multiples (i.e., a series of similar figures with the same scale and axes) can allow a transparent and simple comparison of separately sampled groups (Fig. 3g). As shown in Fig. 3, improving data transparency does not inherently need to make a simple figure complex and with the appropriate application of figure elements to place emphasis, transparent figures can still reduce the need for excessive System 2 cognition.

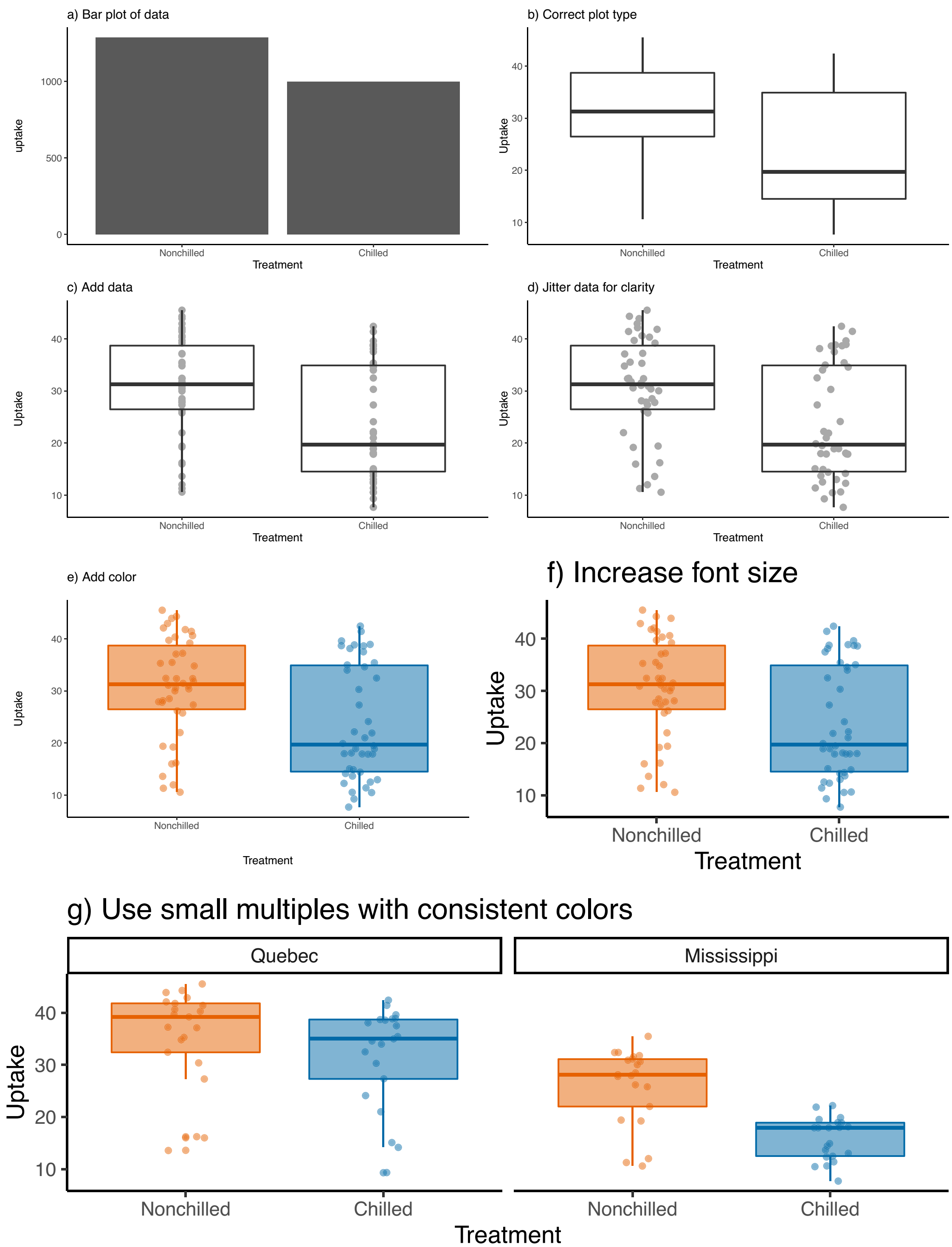

Fig. 3. Example sequence of steps using figure elements and aesthetics to improve the visualization of data (**a**) by using a more informative plot type (**b**), adding the data to the plot (**c**), jittering the data (**d**), adding color (**e**), increasing the font size (**f**), and using small multiples (**g**). The data displayed here are freely available in the CO_2 dataset in the R library datasets and was originally published by Potvin et al. (1990). (All code for producing figures can be found at https://github.com/MatthewRobertson2452/Show-and-Tell) [Correction added on November 3, 2022, after first online publication: The typesetter mistakenly edited the text size within Figure 3 to be standard throughout image. This has now been corrected.]

the underlying data, representing data along meaningful and consistent axis ranges, and ensuring that figure element differences are easily discernable.

Conclusion

Studies and best practices on visual cognition remain cryptic to many figure makers outside of the cognitive sciences. What is more prevalent and accessible are the numerous (and sometimes discipline-specific) articles providing figure-making checklists and other technical suggestions. Here, we try to emphasize both the behind-the-scenes cognitive aspects and the on-stage application—the most effective figures are those that incorporate the importance of both. An apt comparison can be made to writing. Effective writing both features the story arc and plot, while also executing the story through correct grammar, sentence construction, and so on. A great story can fall flat without an ability to effectively put the right words into logical sentences, and a great technical writer may not be successful because they do not have a compelling story. The same goes for figures—visual information risks not being transmitted if visual elements are not used effectively, but simply being able to effectively construct a figure doesn't mean you have relevant information to share.

The pace of scientific advances is greatly influenced by our ability to share results and information with other scientists. Important discoveries may remain unknown if not for an effective means of communication. Unfortunately, those of us unfamiliar with cognitive science are often unaware of small visual considerations that may provide a big boost to visual information sharing. Fortunately, many of the lessons can be learned from cognitive science and are intuitive and easy to implement once they are known. Ultimately, making effective figures is a continuous process of learning, and rarely is there one correct version of any given visualization. Despite this lack of a clear endpoint, we think that some of the ideas presented here will help many practicing scientists to effectively share their message so that the flow of information and ideas can operate with minimal barriers.

References

Anderson, E. W., K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva. 2011. A user study of visualization effectiveness using EEG and cognitive load, p. 791–800. *In Computer graphics forum*, v. **30**, no. 3. Blackwell Publishing Ltd. doi: 10.1111/j.1467-8659.2011.01928.x

Boote, S. K. 2014. Assessing and understanding line graph interpretations using a scoring rubric of organized cited factors. J. Sci. Teacher Educ. **25**: 333–354.

Boote, S. K., and D. N. Boote. 2017. Leaping from discrete to continuous independent variables: Sixth graders' science line graph interpretations. Element. School J. **117**: 455–484. doi:10.1086/690204

Borkin, M. A., Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. 2015. Beyond memorability: Visualization recognition and recall. IEEE Trans. Vis. Comput. Graph. **22**: 519–528.

Correll, M., E. Bertini, and S. Franconeri. 2020. Truncating the y-axis: Threat or menace?. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–12).

Dimara, E., S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. 2018. A task-based taxonomy of cognitive biases for information visualization. IEEE Trans. Vis. Comput. Graph. **26**: 1413–1432.

Evans, J. S. B., and K. E. Stanovich. 2013. Dual-process theories of higher cognition: Advancing the debate. Perspect. Psychol. Sci. **8**: 223–241. doi:10.1177/1745691612460685

Friel, S. N., F. R. Curcio, and G. W. Bright. 2001. Making sense of graphs: Critical factors influencing comprehension and instructional implications. J. Res. Math. Educ. **32**: 124–158. doi:10.2307/749671

Galesic, M., and R. Garcia-Retamero. 2011. Graph literacy: A cross-cultural comparison. Med. Decis. Making **31**: 444457. doi:10.1177/0272989X10373805

Greenacre, M. 2016. Data reporting and visualization in ecology. Polar Biol. **39**: 2189–2205. doi:10.1007/s00300-016-2047-2

Haroz, S., and D. Whitney. 2012. How capacity limits of attention influence information visualization effectiveness. IEEE Trans. Vis. Comput. Graph. **18**: 2402–2410. doi:10.1109/TVCG.2012.233

Hegarty, M. 2011. The cognitive science of visual-spatial displays: Implications for design. Topics Cogn. Sci. **3**: 446–474. doi:10.1111/j.1756-8765.2011.01150.x

Kahneman, D. 2011, *Thinking, fast and slow*. Macmillan.

Kelleher, C., and T. Wagener. 2011. Ten guidelines for effective data visualization in scientific publications. Environ. Model. Software **26**: 822–827. doi:10.1016/j.envsoft.2010.12.006

Mairena, A., M. Dechant, C. Gutwin, and A. Cockburn. 2020. Baseline study of emphasis effects information visualization. Graphics Interface 2020 Conference. https://openreview.net/forum?id=RQaTpcuiJF

Midway, S. R. 2020. Principles of effective data visualization. Patterns. **1**: 100141. doi:10.1016/j.patter.2020.100141

Okan, Y., R. Garcia-Retamero, M. Galesic, and E. T. Cokely. 2012. When higher bars are not larger quantities: On individual differences in the use of spatial information in graph comprehension. Spatial Cogn. Comput. **12**: 195–218.

Okan, Y., R. Garcia-Retamero, E. T. Cokely, and A. Maldonado. 2018. Biasing and debiasing health decisions with bar graphs: Costs and benefits of graph literacy. Q. J. Exp. Psychol. **71**: 2506–2519. doi:10.1177/1747021817744546

Padilla, L. M., S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. 2018. Decision making with visualizations: A cognitive framework across disciplines. Cogn. Res. Princ. Implic. **3**: 1–25.

Potvin, C., M. J. Lechowicz, and S. Tardif. 1990. The statistical analysis of ecophysiological response curves obtained from experiments involving repeated measures. Ecology **71**: 1389–1400. doi:10.2307/1938276

Rolandi, M., K. Cheng, and S. Pérez-Kriz. 2011. A brief guide to designing effective figures for the scientific paper. Adv. Mater. **23**: 4343–4346. doi:10.1002/adma.201102518

Rougier, N. P., M. Droettboom, and P. E. Bourne. 2014. Ten simple rules for better figures. PLoS Comput. Biol. **10**: e1003833. doi:10.1371/journal.pcbi.1003833

Smart, S., and D. A. Szafir. 2019. Measuring the separability of shape, size, and color in scatterplots, p. 1–14. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://dl.acm.org/doi/10.1145/3290605.3300899

Stone, E. R., J. F. Yates, and A. M. Parker. 1997. Effects of numerical and graphical displays on professed risk-taking behavior. J. Exp. Psychol.-Appl. **3**: 243. doi:10.1037/1076-898X.3.4.243

Tufte, E. R. 1983, *The visual display of quantitative information*. Graphic Press.

Vessey, I. 1991. Cognitive fit: A theory-based analysis of the graphs versus tables literature. Decision Sciences **22**: 219–240. doi:10.1111/j.1540-5915.1991.tb00344.x

Vessey, I. 2006. The theory of cognitive fit. Chapter 8, p. 141–183. *In* P. Zhang and D. Galletta [eds.]. Human–computer interaction and management information systems: Foundations.

Wang, Z. H., S. Wei, W. Ding, X. Chen, X. Wang, and K. Hu. 2012. Students' cognitive reasoning of graphs: Characteristics and progression. Int. J. Sci. Educ. **34**: 2015–2041.

Weissgerber, T. L., S. J. Winham, E. P. Heinzen, J. S. Milin-Lazovic, O. Garcia-Valencia, Z. Bukumiric, M. D. Savic, V. D. Garovic, and N. M. Milic. 2019. Reveal, don't conceal: Transforming data visualization to improve transparency. Circulation **140**: 1506–1518. doi:10.1161/CIRCULATIONAHA.118.037777

Acknowledgments

The authors thank Jim Cloern and Patricia Soranno for their support of the development of this article.

Conflict of interest

None declared.

Submitted 29 July 2021
Revised 27 September 2022
Accepted 07 October 2022