Merging Graphics and Text to Better Convey Experimental Results: Designing an "Enhanced Bar Graph"

William D. Berry, *Florida State University*
Matthew Hauenstein, *Florida State University*

ABSTRACT We propose a format for presenting experimental results that combines a graph's strength in facilitating general-pattern recognition with a table's strength in displaying numerical results. The format supplements a conventional bar graph with additional text labels and graphics but also can be based on a dot plot. The resulting enhanced bar graph conveys general patterns about treatment effects; displays point estimates and confidence intervals for all key quantities of interest relevant to testing hypotheses (e.g., first differences in the mean of the dependent variable); and clarifies the interpretation of these quantities as treatment effects. Presenting information in a single figure avoids the need to devote scarce journal space to both a graph and a table. Moreover, an enhanced bar graph prevents readers from having to move back and forth between a graph and a table of numerical results—thereby reducing their cognitive load and facilitating their understanding of the findings.

In the last two decades, randomized experiments have become more common in political science (Druckman et al. 2006).[1] This trend magnifies the importance of clear communication of the statistical results of experiments—and their interpretation as treatment effects—to readers. However, no consensus has emerged among experimentalists about the best format for presenting estimated treatment effects.[2] This diversity in presentational formats is not surprising given that graphs (e.g., bar graphs or dot plots) and tables are widely perceived to have different strengths as vehicles for displaying findings: graphs better convey general patterns, whereas tables are superior for looking up detailed results (Gelman, Pasarica, and Dodhia 2002; Kastellec and Leoni 2007; Lane and Sándor 2009). This conventional wisdom suggests that the best way to communicate experimental findings is to present both a graph and a table, thereby avoiding the need for readers to sacrifice either an ability to quickly discern general patterns or access to specific numerical results.

However, manuscript-length limitations imposed by journals create a disincentive for authors to present evidence about treatment effect in both a graph and a table; indeed, it is rare for published work to present both.[3] Thus, it is valuable to consider: Is it possible to design a "grable"[4] that supplements a graph with the numerical results typically displayed in a table but does not take significantly more space than would be required for the graph alone? We believe not only that the answer is "yes," but that a well-designed grable can convey experimental findings *better* than a combination of separate graph and table. We are convinced by Sweller et al. (1990; see also Chandler and Sweller 1992) that the overriding consideration when presenting information is to minimize a reader's "cognitive load"—that is, the amount of mental processing required to understand the information. Moreover, there is both strong theory (Gillan et al. 1998; Lane and Sándor 2009; Wainer 1997, ch. 17) and experimental evidence (Chandler and Sweller 1992, 178; see also Sweller et al. 1990) that requiring readers to "split their attention between multiple sources of information" (e.g., a graph and a table) imposes a higher cognitive load than consolidating all information in a single display.

Accordingly, we contend that for many experiments, the best strategy for presenting results is to construct a single figure combining graphics, numbers, and text. This grable would (1) rely on graphics to convey general patterns about treatment effects, (2) display specific values for the key quantities of interest that provide detail about the strength and importance of these treatment

William D. Berry *is the Syde P. Deeb Eminent Scholar and Marian D. Irish Professor of Political Science at Florida State University. He can be reached at wberry@fsu.edu.*
Matthew Hauenstein *is a doctoral candidate in the department of political science at Florida State University. He can be reached at mhauenstein@fsu.edu.*

doi:10.1017/S1049096517000683

effects, and (3) use text labels to help readers interpret these quantities. The key to the success of this strategy is in developing a format for a grable that integrates all of this information without overwhelming readers. We believe such integration is feasible because two features of most experimental research combine to keep the number of quantities of interest relatively low: (1) the number of independent variables observed tends to be small, and (2) each independent variable typically has only a small number of discrete values.

Among experimentalists relying on graphs to convey results, a bar graph that shows the mean value of the dependent variable in each experimental condition is a frequently used format.[5] A bar graph is a good choice of graphical format because research shows that readers can successfully judge the length of multiple objects (e.g., bars) plotted alongside an axis depicting a linear scale (Gillan et al. 1998; Jacoby and Schneider 2010; Kosslyn 2006). Yet, the success of a conventional bar graph in facilitating general-pattern recognition is limited by the fact that its bars focus a reader's attention on the mean value of the dependent variable in each experimental condition rather than on the actual quantities of interest in an experiment. These quantities nearly always include first differences in means across experimental conditions (which capture the strength of treatment effects). When the theory being tested posits interaction between independent variables, the quantities also typically include "differences between differences" (or second differences) that reflect variation in treatment effects across contexts.

quantitative methods. We recognize that some scholars argue that dot plots are superior to bar graphs for communicating statistical results (e.g., Cleveland 1984; Jacoby 2006). Accordingly, we illustrate how a dot plot can be "enhanced" to convey the same information as an enhanced bar graph.

KEY QUANTITIES OF INTEREST FROM AN EXPERIMENT

For most political science experiments, a well-designed table can display all relevant quantities of interest. To illustrate, consider the results of a fictitious 2x2 factorial experiment to test a hypothesis that two variables—Treatment 1 (absent or present) and Treatment 2 (absent or present)—interact in influencing a dependent variable, Y. The hypothesis is that each treatment increases Y regardless of whether the other treatment is present, and each treatment is more effective when the other is present than when the other is absent. Table 1 reports (in the nonshaded cells) the mean value of Y—to be denoted $\bar{Y}$—in each of the four experimental conditions (along with the sample size for the condition). The table also displays point estimates of all quantities of interest relevant to testing the hypothesis: (1) in the four lightly shaded cells, first differences in $\bar{Y}$ reflecting the average effect of each treatment when the other treatment is present as well as when the other treatment is absent; and (2) in the darkly shaded cell, the second difference (or "difference between differences") in $\bar{Y}$, capturing the strength of interaction between Treatment 1 and Treatment 2. Finally, table 1 presents a 95% confidence interval for each estimated quantity.

The key to the success of this strategy is in developing a format for a grable that integrates all of this information without overwhelming readers.

We propose supplementing a conventional bar graph—with additional text and graphics—to direct a reader's attention to these quantities of interest. We claim that for a typical experiment, the resulting enhanced bar graph (1) conveys general patterns in the results better than a standard bar graph; (2) displays all key quantities of interest relevant to testing hypotheses—both point estimates and confidence intervals—in locations that are easy for a reader to find; and (3) clarifies how these quantities can be interpreted as estimated treatment effects. The third feature is especially valuable to readers with limited training in

Table 1 provides "look-up" capability because readers can find within the table each of the five quantities of interest (i.e., four first differences and one second difference). However, for a reader not already familiar with the table's format or not well trained in experimental design, discerning the strength of treatment effects and the extent of interaction from table 1 requires careful inspection. This leads us to consider: Can we design a figure that would allow a reader—even one rarely exposed to experimental research—to easily discern the numeric value of each quantity of interest in table 1, yet also display a graph that makes immediately

Table 1

Results from a Fictitious 2x2 Factorial Experiment to Test a Hypothesis that the Two Factors Interact in Influencing a Dependent Variable

		(1)	(2)	(3)
		Treatment 2 absent	**Treatment 2 present**	**Column 2 minus Column 1**
(A)	Treatment 1 absent	**111.7** (109.3, 114.1) n = 20	**127.0** (124.6, 129.4) n = 20	**15.3** (11.9, 18.7)
(B)	Treatment 1 present	**114.4** (112.0, 116.7) n = 20	**136.2** (133.8, 138.5) n = 20	**21.8** (18.4,25.2)
(C)	**Row B minus row A**	**2.6** (-0.7, 6.0)	**9.1** (5.8,12.5)	**6.5** (1.7, 11.3)

Note: Each of the four nonshaded cells of the table shows the estimated mean of the dependent variable, Y, among subjects in an experimental condition. Each row or column marginal (i.e., lightly shaded cell) reports a difference in means reflecting the effect of one factor at a value of the other factor. The darkly shaded cell in the lower-right corner contains the difference between two differences in means and reflects the strength of interaction between the two factors in their effect on Y. Each estimated quantity is reported along with the boundaries for a 95% confidence interval in parentheses.

https://doi.org/10.1017/S1049096517000683 Published online by Cambridge University Press

evident the experiment's general conclusions about the hypothesized treatment effects?

ENHANCING A CONVENTIONAL BAR GRAPH

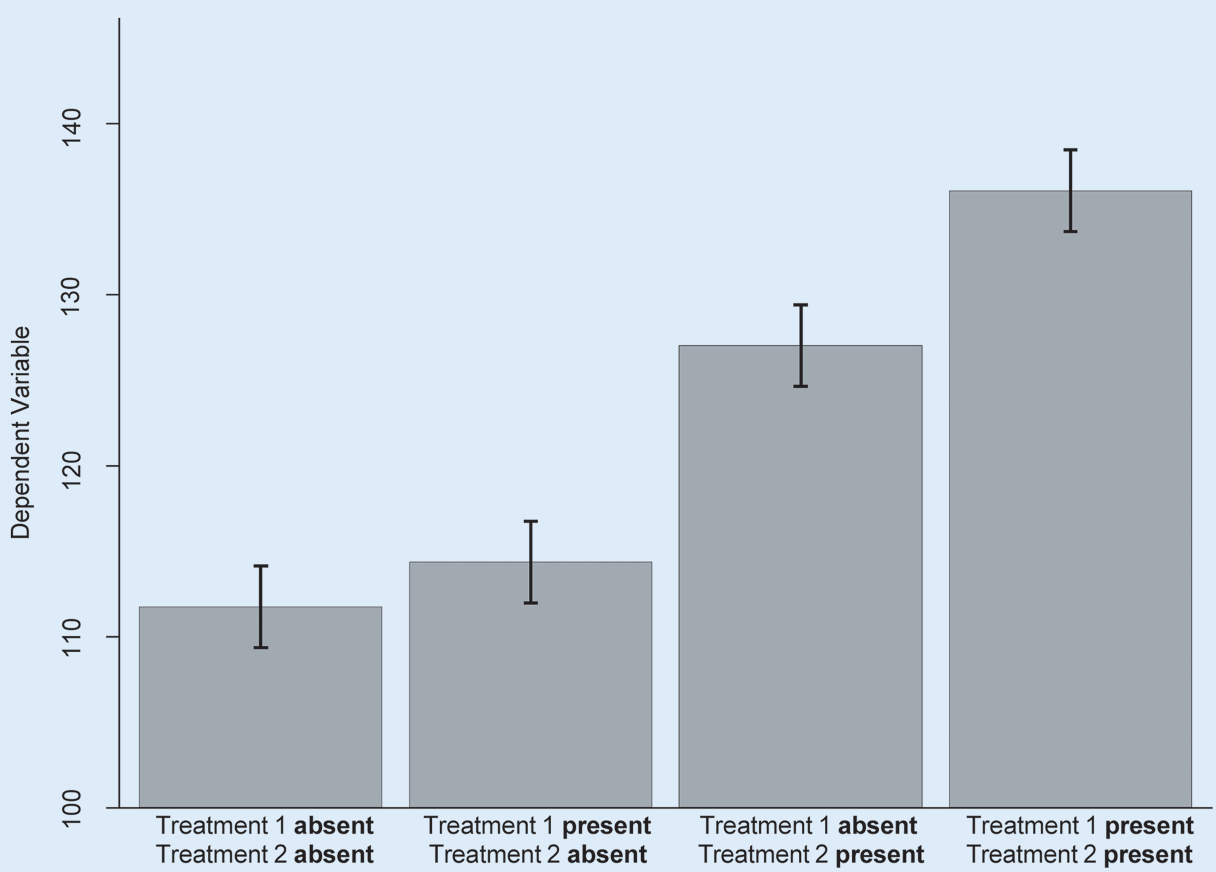
Figure 1 portrays the results from table 1 in a conventional bar graph; that is, the graph plots the estimated mean of *Y*—along with a 95% confidence interval—in each of the four experimental conditions. Figure 1 clearly outperforms table 1 in facilitating pattern recognition. Ignoring the strength of treatment effects and considering only their direction, we can easily compare (1) the lengths of the first and second bars to discern that Treatment 1's effect is positive in the absence of Treatment 2; and (2) the lengths of the third and fourth bars to recognize that Treatment 1's effect is positive in the presence of Treatment 2. We also can observe that the effect of Treatment 2 is positive regardless of whether Treatment 1 is present. However, this recognition requires a more demanding task: comparing the lengths of two nonadjacent bars (i.e., the first to the third, and the second to the fourth). Finally, we can see that Treatment 1's effect is stronger when Treatment 2 is present by recognizing that the difference in the lengths of the right-most two bars (reflecting the effect of Treatment 1 in the *presence* of Treatment 2) is greater than the difference in the lengths of the left-most two bars (reflecting the effect of Treatment 1 in the *absence* of Treatment 2). This is clearly the most challenging of the pattern recognitions because it requires simultaneous consideration of the lengths of all four bars. Readers specializing in experimental research are likely to recognize quickly that the pattern of the four bars is indicative of interaction between Treatment 1 and Treatment 2; however, those less familiar with experiments may need to more closely examine the graph to see that the interaction is present.

Thus, there is room to improve a conventional bar graph's ability to facilitate general-pattern recognition. The bars in a conventional bar graph steer a reader's attention to the value of $\bar{Y}$ in each experimental condition. With additional graphics, attention can be directed instead to the quantities of interest relevant to testing a researcher's hypotheses: the first and second differences in $\bar{Y}$. Furthermore, the conventional bar graph in figure 1 depicts confidence intervals only for $\bar{Y}$ values and provides no information about the uncertainty of the estimated quantities of interest—which is far more important.

We overcome the deficiencies of a conventional bar graph with additional text and graphics to produce an enhanced bar graph. The online appendix contains a detailed description of the features of an enhanced bar graph. This article illustrates these features by supplementing figure 1 with additional information to produce the enhanced bar graph in figure 2.[6] Figure 2 conveys all relevant information about each of the

Figure 1

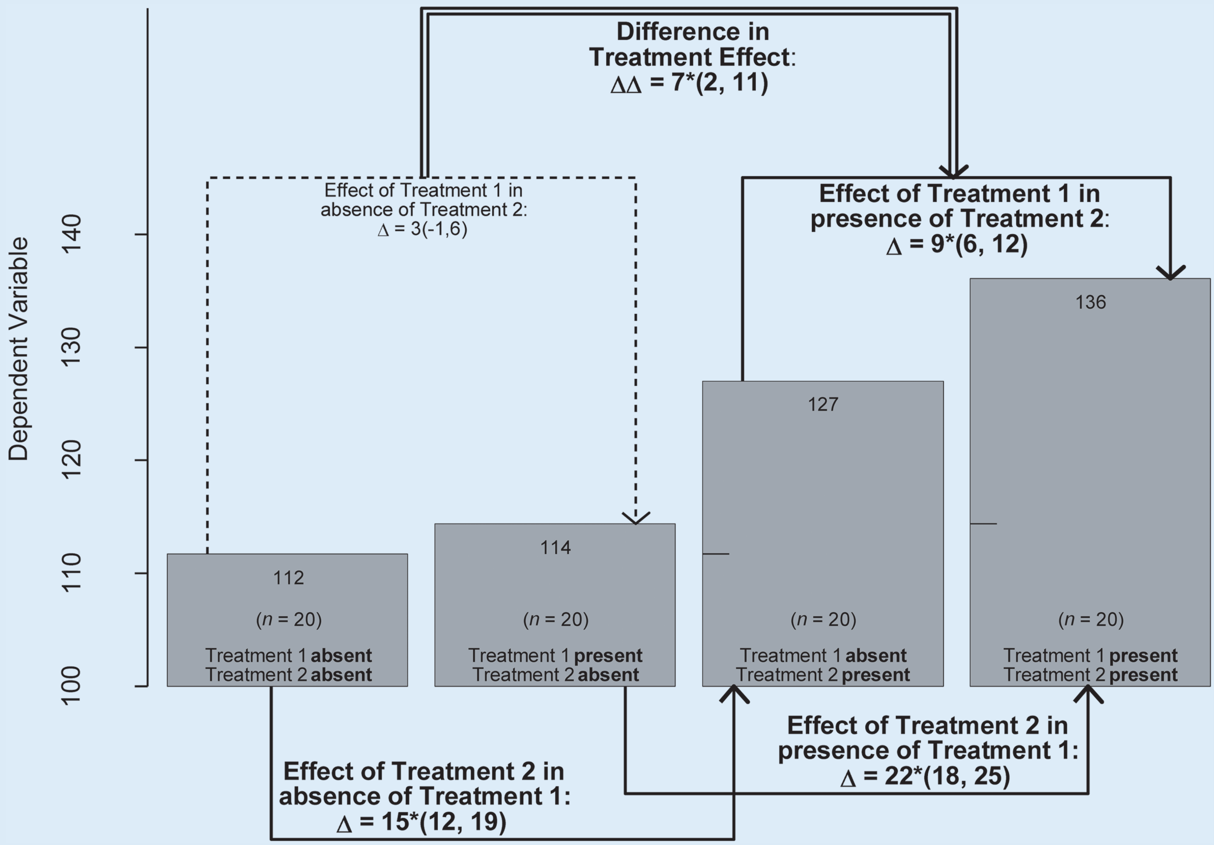
A Conventional Bar Graph Showing Results from Table 1



Note: Each bar shows the estimated mean of the dependent variable among subjects in an experimental condition. The vertical line overlaid on a bar shows the boundaries for a 95% confidence interval.

Figure 2

An Enhanced Bar Graph Showing Results from Table 1



Note: Each bar shows the estimated mean of the dependent variable among subjects in an experimental condition. Each Δ value next to a U-shaped arrow is a first difference in means reflecting a treatment effect; a solid arrow indicates an effect deemed substantively significant. The ΔΔ value next to the double arrow is a second difference (i.e., a difference between two differences in means: 7 ≈ 9 – 3) reflecting the strength of interaction. (ΔΔ could be computed equivalently as a difference between the other two first differences portrayed: 7 = 22 – 15.) Each estimated quantity of interest is reported along with the boundaries for a 95% confidence interval in parentheses. An asterisk (*) indicates statistical significance at the 0.05 level (two-tailed test).

five quantities of interest necessary to evaluate the underlying hypothesis without requiring readers to alternate between a graph and a table of numerical results. The figure relies on the following several conventions:

- U-shaped arrows are included to focus a reader's attention on the key quantities of interest: first and second differences in means (i.e., in the lengths of bars). Each first difference in means is depicted with a single arrow connecting the two experimental conditions being compared. The second difference is portrayed with a double arrow connecting the two first differences being compared. An arrow is made solid to indicate that a difference is statistically significant and deemed large enough in magnitude to be substantively important. An arrow is dashed to convey that a difference is statistically insignificant or too small to be of practical consequence.
- Text labels are used to display each numerical quantity of interest, with the symbol Δ denoting a first difference and the symbol ΔΔ indicating a second difference. We use large, boldface text to indicate quantities that are statistically significant and substantively important; smaller, lightface text is used for quantities that are statistically or substantively insignificant. We also choose a level of precision for each quantity that avoids displaying substantively trivial digits that serve only to distract.
- The text label for each quantity of interest provides details relevant to determining both its statistical and practical significance: a point estimate, followed by a 95% confidence interval. However, to avoid unnecessary clutter, we deviate from the conventional bar graph in figure 1 by *not* displaying a confidence interval for the mean value of the dependent variable in each of the four experimental conditions; these means are not relevant to testing the hypothesis underlying the experiment.
- The point estimate of each quantity of interest is preceded by an interpretation of its meaning (e.g., in the upper-right corner of figure 2, "Effect of Treatment 1 in presence of Treatment 2"). Of course, in a well-written paper, such interpretations are thoroughly discussed in the text. However, we believe that incorporating brief descriptions of interpretations of relevant first and second differences can help readers recognize how the interpretations emerge from the statistical results. This feature of an enhanced bar graph is especially valuable for those without strong training in quantitative methods.

https://doi.org/10.1017/S1049096517000683 Published online by Cambridge University Press

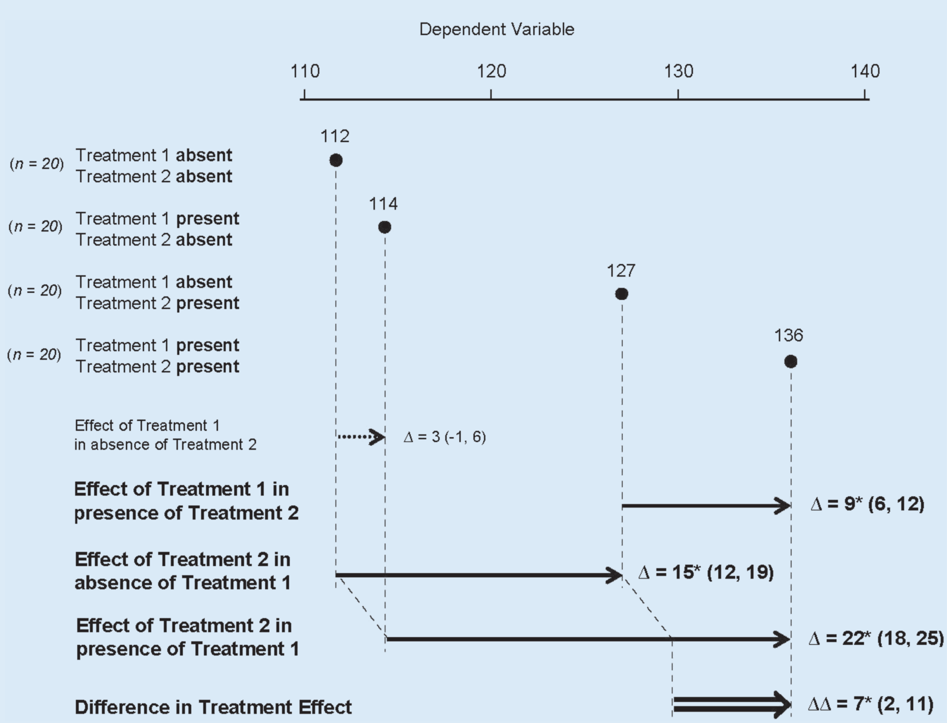
In our view, incorporating the arrows and text labels into the bar graph in figure 2 does not detract from a reader's ability to quickly absorb the general patterns evident by scanning the relative lengths of bars. Moreover, by strategically positioning the text in the enhanced bar graph, we can guide readers to relevant quantities of interest, thereby enhancing their ability to connect these quantities to the general patterns in a way that facilitates understanding of the experimental results. Many experiments in political science have four or fewer observed experimental conditions; for these studies, we believe that an easily readable enhanced bar graph can almost always be constructed.[7]

…by strategically positioning the text in the enhanced bar graph, we can guide readers to relevant quantities of interest, thereby enhancing their ability to connect these quantities to the general patterns in a way that facilitates understanding of the experimental results.

Figure 3 depicts a version of a dot plot supplemented with numerical values and text to convey the same information presented in the enhanced bar graph shown in figure 2. At the top of the enhanced dot plot, there are four dots indicating the mean of *Y* in each experimental condition. Below these dots are five arrows. Each single arrow represents a first difference in means; the double arrow denotes the second difference. We believe that the principal advantage of the dot plot over the bar graph is that the former displays each quantity of interest using an object—an arrow—with a length equal to the quantity. However, this advantage is lessened by the fact that the arrows are not aligned to start at the same origin, which complicates visual comparison of the magnitude of treatment effects. As a consequence, we believe it is easier for readers to see how each first and second difference is computed from component $\bar{Y}$ values in the bar graph than in the dot plot. Balancing all considerations, we think an enhanced bar graph is a slightly more effective grable than an enhanced dot plot for conveying experimental findings.[8]

Figure 3

An Enhanced Dot Plot Showing the Same Results as Figure 2 Using Straight Arrows To Depict First and Second Differences.



Note: Each dot shows the estimated mean of the dependent variable among subjects in an experimental condition. Each Δ value to the right of an arrow is a first difference in means reflecting a treatment effect; a solid arrow indicates an effect deemed substantively significant. The ΔΔ value to the right of the double arrow is a second difference (i.e., a difference between two differences in means: 7 = 22 − 15) reflecting the strength of interaction. (ΔΔ could be computed equivalently as a difference between the other two first differences portrayed: 7 ≈ 9 − 3.) Each estimated quantity of interest is reported along with the boundaries for a 95% confidence interval in parentheses. An asterisk (*) indicates statistical significance at the 0.05 level (two-tailed test).

Note: An online appendix and code in both Stata and R illustrating how an enhanced bar graph can be constructed are available at coss.fsu.edu/enhancedbargraph. We encourage researchers who construct enhanced bar graphs to present their experimental results to share their computer code with other scholars. To facilitate this sharing, if researchers e-mail their code along with a pdf image of the graph created to one of us, we will post the files at the website, explicitly recognizing the generosity of the contributor.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S1049096517000683

ACKNOWLEDGMENTS

We thank Kevin Dyrland for assistance in creating the graphs portrayed in the article; and Kevin Arceneaux, Jason Barabas, Jamie Druckman, Jens Grosser, Jennifer Jerit, Thomas Leeper, Kevin Mullinix, Megan Shannon, and Mark Souva for helpful comments on previous versions of it. We are also grateful to Vera Mironova and Sam Whitt for sharing replication data. ■

NOTES

1. This trend is underscored by the founding of an organized APSA section on experimental research in 2010 and the publication of the first issue of *Journal of Experimental Political Science* (*JEPS*) in 2014.
2. For example, 11 articles published in the two inaugural 2014 issues of *JEPS* reported at least one estimated treatment effect. When reporting these estimates, six of the articles relied exclusively on tables, two relied solely on figures, two used a combination of figures and tables, and one used neither figures nor tables (relying only on text). See table A-1 in the online appendix for details.
3. See note 2.
4. The earliest usage we can find of the term *grable*—to describe a combination graph/table—is by Hink, Wogalter, and Eustace (1996).
5. Three articles in the two 2014 issues of *JEPS* used one or more figures to depict the mean value of the dependent variable in each experimental condition; two used bar graphs (Broockman 2014; Stadelmann, Portmann, and Eichenberger 2014); and one used a dot plot (Healy, Kuo, and Malhotra 2014).
6. The online appendix contains two other examples of an enhanced bar graph: one conveying the results from a one-factor experiment to test a hypothesis that *Y* is greater in the presence of a treatment than in its absence (see figure A-3), the other depicting findings from a one-factor (four-level) experiment to test a hypothesis that each increase in the level of a treatment produces an increase in *Y* (see figure A-4).
7. Indeed, the findings of some studies with as many as six experimental conditions can be effectively conveyed using an enhanced bar graph with landscape orientation. For example, there are three articles in the two 2014 issues of *JEPS* presenting experiments that involve six conditions for which we think an enhanced bar graph would be a good format: by Mironova and Whitt (2014, table 1) (with six first differences to be displayed); and Al-Ubaydli, McCabe, and Twieg (2014, table 1) and Krupnikov and Levine (2014, table 3) (each of which would display three first differences and three second differences).
8. Figure A-9 in the online appendix presents the same results as figure 2 using an alternative display format that involves text boxes but no graphical elements. It sacrifices the pattern-clarifying advantages of graphs. However, it shows that the difference in treatment effects displayed can be computed in two different ways.

REFERENCES

Al-Ubaydli, Omar, Kevin McCabe, and Peter Twieg. 2014. "Can More Be Less? An Experimental Test of the Resource Curse." *Journal of Experimental Political Science* 1: 39–58.

Broockman, David E. 2014. "Mobilizing Candidates: Political Actors Strategically Shape the Candidate Pool with Personal Appeals." *Journal of Experimental Political Science* 1: 104–19.

Chandler, Paul and John Sweller. 1992. "The Split-Attention Effect as a Factor in the Design of Instruction." *British Journal of Educational Psychology* 62: 233–46.

Cleveland, William S. 1984. "Graphical Methods for Data Presentation: Full-Scale Breaks, Dot Charts, and Multibased Logging." *The American Statistician* 38: 270–80.

Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia. 2006. "The Growth and Development of Experimental Research in Political Science." *American Political Science Review* 100: 627–36.

Gelman, Andrew, Cristian Pasarica, and Rahul Dodhia. 2002. "Let's Practice What We Preach: Turning Tables into Graphs." *The American Statistician* 56: 121–30.

Gillan, Douglas J., Christopher D. Wickens, J. G. Hollands, and C. Melody Carswell. 1998. "Guidelines for Presenting Quantitative Data in HFES Publications." *Human Factors* 40: 28–41.

Healy, Andrew, Alexander G. Kuo, and Neil Malhotra. 2014. "Partisan Bias in Blame Attribution: When Does It Occur?" *Journal of Experimental Political Science* 1: 144–58.

Hink, Jessica K., Michael S. Wogalter, and Jason K. Eustace. 1996. "Display of Quantitative Information: Are Grables Better than Plain Graphs or Tables?" *Proceedings of the Human Factors and Ergonomics Society*, 40th Annual Meeting, 1155–9.

Jacoby, William G. 2006. "The Dot Plot: A Graphical Display for Labeled Quantitative Values." *The Political Methodologist* 14: 6–14.

Jacoby, William G. and Saundra Schneider. 2010. "Graphical Displays for Political Science Journal Articles." Paper presented at the Visions in Methodology Conference, Iowa City, IA, March.

Kastellec, Jonathan P. and Eduardo L. Leoni. 2007. "Using Graphs Instead of Tables in Political Science." *Perspectives on Politics* 5: 755–71.

Kosslyn, Stephen Michael. 2006. *Graph Design for the Eye and Mind*. Cary, NC: Oxford University Press.

Krupnikov, Yanna and Adam Seth Levine. 2014. "Cross-Sample Comparisons and External Validity." *Journal of Experimental Political Science* 1: 59–80.

Lane, David M. and Anikó Sándor. 2009. "Designing Better Graphs by Including Distributional Information and Integrating Words, Numbers, and Images." *Psychological Methods* 14: 239–57.

Mironova, Vera and Sam Whitt. 2014. "Ethnicity and Altruism after Violence: The Contact Hypothesis in Kosovo." *Journal of Experimental Political Science* 1: 170–80.

Stadelmann, David, Marco Portmann, and Reiner Eichenberger. 2014. "Full Transparency of Politicians' Actions Does Not Increase the Quality of Political Representation." *Journal of Experimental Political Science* 1: 16–23.

Sweller, John, Paul Chandler, Paul Tierney, and Martin Cooper. 1990. "Cognitive Load as a Factor in the Structuring of Technical Material." *Journal of Experimental Psychology: General* 119: 176–92.

Wainer, Howard. 1997. "Improving Tabular Displays, with NAEP Tables as Examples and Inspirations." *Journal of Educational and Behavioral Statistics* 22: 1–30.

https://doi.org/10.1017/S1049096517000683 Published online by Cambridge University Press