

Online Appendix

for

Merging Graphics and Text to Better Convey Experimental Results: Designing an “Enhanced Bar Graph”*

William D. Berry
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 can also 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—reducing their cognitive load, and facilitating their understanding of the findings.

IMPORTANT NOTE TO READERS: So that this online appendix is comprehensible if you read it as a stand-alone document without having first read the associated article in *PS*, most of the text of the article is incorporated verbatim into the first six pages of the appendix. The few portions of the first five pages of the appendix that are not drawn verbatim from the *PS* article appear in **red text** below. **Accordingly, if you are reading this appendix after having read the *PS* article, we recommend that you first read the red text on pages 3 and 4. Then you should skip to page 5 and read the rest of the appendix starting with the paragraph beginning, “Below we elaborate...”**

*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 the sharing, if someone emails one of us his or her code along with a pdf image of the graph created, we will post the files at the website, explicitly recognizing the generosity of the contributor.**

We thank Kevin Dyrland for assistance in creating the graphs portrayed in the paper; 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 the paper. We are also grateful to Vera Mironova and Sam Whitt for sharing replication data.

**Table of Contents for Numbered Figures, Tables and Documents
in the Online Appendix**

Note: Table 1 and Figures 1, 2 and 3 are taken directly from the *PS* article.

Table 1	15
Figure 1	16
Figure 2	17
Figure 3	18
Document A-1	19
Figure A-1	20
Figure A-2	21
Figure A-3	22
Figure A-4	23
Figure A-5	24
Figure A-6	25
Figure A-7	26
Figure A-8	27
Figure A-9	28
Table A-1	29

Over the last two decades, randomized experiments have become much 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, while tables are superior for looking up detailed results (Gelman, Pasarica and Dodhia 2002; Kastlelec 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 a 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 a treatment effect in both a graph and a table; and, 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,” i.e., the amount of mental processing required to understand the information. And there is both strong theory (Lane and Sándor 2009; Wainer 1997, ch. 17; Gillan et al. 1998) 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 on readers 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 (i) rely on graphics to convey general patterns about treatment effects, (ii) display specific values for the key quantities of interest that provide detail about the strength and importance of these treatment effects, and (iii) use text labels to help readers interpret these quantities. The key to the success of this strategy is developing a format for a grable that integrates all this information without overwhelming the reader. We believe such integration is feasible because two features of most experimental research combine to keep the number of quantities of interest relatively

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, eleven articles published in the two inaugural 2014 issues of *JEPS* report at least one estimated treatment effect. When reporting these estimates, six of the articles rely exclusively on tables, two rely solely on figures, two use a combination of figures and tables, and one uses neither figures nor tables (relying only on text); see Table A-1 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).

low: the number of independent variables observed tends to be small, and 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 very successfully judge the length of multiple objects (e.g., bars) plotted alongside an axis depicting a linear scale (Jacoby and Schneider 2010; Kosslyn 2006; Gillan et al. 1998). Yet the success of a conventional bar graph at 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 the actual quantities of interest of an experiment. These quantities nearly always include first differences in means across experimental conditions (which capture the strength of treatment effects); and 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.

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* (i) conveys general patterns in the results better than a standard bar graph, (ii) 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 (iii) clarifies how these quantities can be interpreted as estimated treatment effects. The last feature is especially valuable to readers with limited training in quantitative methods. We recognize that some scholars have argued that dot plots are superior to bar graphs for communicating statistical results (e.g., Jacoby 2006; Cleveland 1984). Accordingly, we illustrate how a dot plot can be “enhanced” to convey the same information as an enhanced bar graph.

The 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 2 x 2 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 non-shaded 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: (i) in the four lightly-shaded cells, first differences in \bar{Y} reflecting the average effect of each treatment both when the other treatment is present and when the other treatment is absent, and (ii) in the darkly-shaded cell, the second difference (or “difference between differences”) in \bar{Y} , capturing the strength of interaction

⁵ Three articles in the two 2014 issues of *JEPS* use one or more figures to depict the mean value of the dependent variable in each experimental condition; two use bar graphs (Broockman 2014; Stadelmann, Portmann and Eichenberger 2014); one uses a dot plot (Healy, Kuo and Malhotra 2014).

between *Treatment 1* and *Treatment 2*. Finally, the table presents a 95% confidence interval for each estimated quantity.

Table 1 gives a reader “look-up” capability, since one can find within the table each of the five quantities of interest (four first differences, and one second difference).⁶ 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 would require very 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 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; the graph plots the estimated mean of Y —along with a 95% confidence interval—in each of the four experimental conditions. This figure clearly outperforms Table 1 in facilitating pattern recognition. Setting aside the strength of treatment effects and considering only their direction, one can easily compare (i) the lengths of the first and second bars to discern that treatment 1’s effect is positive in the absence of treatment 2, and (ii) the lengths of the third and fourth bars to recognize that treatment 1’s effect is positive in the presence of treatment 2. One can also observe that the effect of treatment 2 is positive regardless of whether treatment 1 is present, yet this recognition requires a somewhat more demanding task—comparing the lengths of two nonadjacent bars (the first to the third, and the second to the fourth). Finally, one 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, as it requires simultaneous consideration of the lengths of all four bars. A reader specializing in experimental research is likely to recognize quickly that the pattern of the four bars is indicative of interaction between *Treatment 1* and *Treatment 2*, but readers less familiar with experiments may need to examine the graph much more closely to see that 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

⁶ Table 1 should be contrasted with a table of raw coefficient estimates that contains the information necessary to generate the quantities of interest displayed in Table 1. For example, one might derive the estimates in Table 1 from a regression of Y on *Treatment 1*, *Treatment 2*, and the product *Treatment 1*Treatment 2*. We believe that a table presenting the coefficient estimates for this regression model—or a dot plot showing this information—is a poor format for conveying experimental results because it does not explicitly present an estimate for each quantity of interest.

intervals only for \bar{Y} values, and offers no information about the uncertainty of the estimated quantities of interest—which is far more important.⁷

We can turn to other academic disciplines for a partial solution to these weaknesses of Figure 1. In particular, bar graphs in psychology journals often alert a reader that a first difference in \bar{Y} is statistically significant by placing a bracket (or a line) above and extending across the two bars associated with the difference, and an asterisk (or a symbol like “ $p < 0.05$ ”) next to the bracket.⁸ Following this bracket/asterisk convention converts Figure 1 into Figure A-1. Although Figure A-1 conveys more information to readers, it fails to recognize (i) the widely understood limitations of a narrow focus on a dichotomous “significant vs. nonsignificant” decision rule, and (ii) the value of presenting a confidence interval for a first difference—which helps readers understand the practical importance of a treatment effect (Gelman and Stern 2006; Krantz 1999). Moreover, the graph in Figure A-1 offers no information about the uncertainty of the second difference in \bar{Y} reflecting the strength of interaction between the two treatments.

We overcome the deficiencies in Figures 1 and A-1 with additional text and graphics—to produce the *enhanced bar graph* in Figure 2. This figure relies on several conventions to convey all relevant information about each of the five quantities of interest necessary to evaluate the underlying hypothesis without requiring a reader to go back and forth between a graph and a table of numerical results:

- 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 $\Delta\Delta$ indicating a second difference. We use large bold text to indicate quantities that are statistically significant and substantively important; and smaller, lighter text 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

⁷ However, Arceneaux and Johnson (2015) note that if the confidence intervals displayed in Figure 1 were at a level of 83.5% rather than 95%, readers could observe whether two intervals overlap to test whether a first difference in \bar{Y} is statistically significant at the .05 level (two-tailed test).

⁸ Indeed, a single recent issue of *Journal of Experimental Psychology* (April 2015) contains three papers using this convention. Our unsystematic review of published articles suggests that psychology may be the only social science discipline in which this convention has wide use. However, we also find it used frequently in the biological (e.g., Aidala et al. 2015; Chatterjee et al. 2012), and medical (e.g., Hirashima et al. 2009; DiPaolo et al. 2008) sciences.

in Figure 1 by *not* displaying a confidence interval for the mean value of the dependent variable in each of the four experimental conditions, since these means are not themselves 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., near 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 paper’s 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 readers without strong training in quantitative methods.

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 from scanning the relative lengths of bars. Moreover, by positioning the text in the enhanced bar graph strategically, we can guide readers to relevant quantities of interest, and thus enhance their ability to connect these quantities to the general patterns in a way that facilitates understanding 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 of 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 are five arrows. Each single arrow represents a first difference in means, and 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—complicating visual comparison of the magnitude of treatment effects. As a consequence, we believe it is easier for a reader 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 vehicle for conveying experimental findings than an enhanced dot plot.⁹

Below we elaborate on the conventions illustrated in Figure 2 for conveying the key quantities of interest from an experiment—and their interpretation—in an enhanced bar graph.¹⁰ Some experiments involve too many distinct experimental conditions to permit presentation of findings in a single simple graph. But many experiments in political science have four or fewer observed experimental conditions, and for such studies, we believe that an easily readable enhanced bar graph can nearly always be constructed.¹¹ We illustrate the features of such a

⁹ Figure A-9 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. But it makes very evident that the difference in treatment effects displayed can be computed in two different ways.

¹⁰ Figure 2 involves four treatment effects ranging from very weak to very strong, and strong interaction between the two treatments. In general, the weaker the treatment effects and the weaker the interaction, the less variation in the length of bars, and thus, the more space left for the U-shaped arrows and text labels. Nevertheless, the conventions illustrated in Figure 2 are applicable regardless of the strengths of treatment effects and the strength of interaction.

¹¹ 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,

graph with figures showing the results from three fictitious experiments: (i) a one-factor experiment to test a hypothesis that Y is greater in the presence of a treatment than in its absence (Figure A-3), (ii) a one-factor (four-level) experiment to test a hypothesis that each increase in the level of a treatment produces an increase in Y (Figure A-4), and (iii) the 2 x 2 factorial experiment described above to test a hypothesis that the two independent variables interact in influencing Y (Figure 2). We begin, however, by delineating several general principles that guide the design of an enhanced bar graph.

Guiding Principles for Designing an Enhanced Bar Graph

Our design of an integrated figure is guided by several principles.

(1) **Rely on familiar conventions.** To facilitate communication of results, when feasible, use display formats to which readers are already accustomed (Kosslyn 2006, 34; Gillan et al. 1998). Indeed, it is this principle that leads us to rely on a bar graph as the basis for the figure we recommend.

(2) **Use graphics to facilitate rapid pattern recognition.** The graphics in a figure should be designed so that a quick perusal makes evident whether the empirical evidence is consistent with the hypotheses being tested (Gillan et al. 1998).

(3) **Use text to enable easy “look up” of each quantity of interest, varying the size and darkness of text to signal the relative importance of information** (Gillan et al. 1998).

(4) **Avoid unnecessary clutter.** No quantity, text, or graphical element should be included in the figure unless it is directly relevant to testing the hypotheses underlying the experiment, or valuable for interpreting experimental results (Gillan et al. 1998).¹²

An Enhanced Bar Graph for the Simplest Experiment: One Factor with a Binary Treatment

Consider a fictitious one-factor experiment designed to test the hypothesis that some treatment increases the value of a dependent variable, Y . Assume that after treatment, the mean value of Y

there are three articles in the two 2014 issues of *JEPS* presenting experiments involving 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]. As an illustration, we include an enhanced bar graph presenting the results from Mironova’s and Whitt’s (2014) Table 1 in our companion document, “Documentation for Stata and R Code to accompany ‘Merging Graphics and Text to Better Convey Experimental Results: Designing an Enhanced Bar Graph,’” which is available at coss.fsu.edu/enhancedbargraph.

¹² This principle is consistent with Tufte’s (1983) advice to “maximize the data-ink ratio.” One type of clutter is reporting too many digits of information for estimated quantities. There are a variety of rules of thumb for determining how many digits to present based on arguments about statistical precision (Cole 2015). Yet, we would avoid slavish conformity to any generic rule in favor of “Goldilocks rounding”—“not too much and not too little” (Cole 2015, 1). The right-most digit presented for a quantity should be the last digit for which the marginal benefit of knowing the digit (for meaningfully discriminating among values for the quantity) is appreciable.

is 10.0 in the treatment group and 5.2 in the control group. The first difference (in means) across the two experimental conditions is 4.8 ($= 10.0 - 5.2$), a value that is statistically significant, and is deemed substantively significant as well. Figure A-2 presents a bar graph displaying these results in a format typical of that found in political science literature.

We believe that the format of this graph can be modified in several ways to better convey the results—yielding the enhanced bar graph in Figure A-3. First, there is plenty of space in Figure A-2 to include the numerical point estimate of the mean of Y in each experimental condition—as we have done in Figure A-3 by displaying this estimate at the right end of the bar for the condition. This eliminates a reader’s need to search for this information elsewhere in the article.

Second, the primary graphical elements in Figure A-2—its two bars—focus the reader’s attention on the magnitude of each of the estimated \bar{Y} values, and not on the key quantity of interest—the difference in means. In the enhanced bar graph in Figure A-3, we add (i) a vertical dotted line demarking the length of the longer bar, (ii) an arrow whose length, 4.8, is clearly equal to the difference between the lengths of the two bars, and (iii) a text label making explicit that this difference is an estimate of the treatment effect.¹³ We rely on this dotted line/arrow convention to capitalize on the fact that people are good judges of the length of an object when it is plotted against an axis with a linear scale (Jacoby and Schneider 2010; Kosslyn 2006).

Third, in Figure A-2, the only information pertaining to statistical uncertainty is about the two \bar{Y} values; there is no information about the uncertainty of the point estimate of the quantity of interest. This problem is overcome in Figure A-3 by (i) including in the text label next to the arrow the 95% confidence interval for the difference in means, and (ii) adding an asterisk to quickly convey that this difference is statistically significant.

Note that this graph conforms to each of the guiding principles delineated above. We rely on familiar conventions (using a bar graph as the core of the figure, and using an asterisk to denote statistical significance). We design the figure’s graphics to facilitate recognition of the most important pattern in the results (the difference between the lengths of the two bars, reflecting the treatment effect), yet simultaneously permit easy look up of the point estimate and confidence interval for the treatment effect. We also avoid unnecessary clutter by abandoning the depiction of confidence intervals for the category means—5.2 and 10.0—which are not quantities directly relevant to testing the hypothesis of interest.

Depicting a difference in means in an enhanced bar graph by including an arrow whose length is this difference is feasible when the two bars corresponding to the difference are adjacent. However, when there are more than two experimental conditions and the quantities of interest include more than one first difference, often some of these differences will involve comparisons of nonadjacent bars. And when two bars involved in a first difference are not adjacent, a dotted line that establishes the location of the arrow depicting the difference may need to cut across a third bar—which would greatly complicate the graphics. Moreover, when interaction between independent variables is hypothesized, typically the quantities of interest will include a second difference (i.e., a “difference between differences”). We see no way to adapt

¹³ Darby and Sloutsky (2015, Figure 8; see also Lane and Sándor 2009, Figure 9) modify a traditional bar graph to display a first difference in means in a slightly different way—by adding a bar for the difference. We believe that using an arrow to denote the difference better clarifies that the estimated treatment effect is equal to this difference.

the dotted line/arrow convention suitable for displaying a first difference across adjacent bars to depict a second difference in a visually appealing way.

When an experiment's design makes the dotted-line/arrow convention illustrated in Figure A-3 insufficient to display each of the study's quantities of interest, we recommend that first and second differences be displayed using U-shaped arrows of the kind illustrated in Figure 2. This sacrifices the graphical benefit of conveying a difference by displaying an object with a length equal to the difference. But strategically positioned U-shaped arrows are a good "second-best" alternative to convey first or second differences, and Figure A-4 illustrates that this vehicle permits reporting as many as six first differences in a single enhanced bar graph without overly cluttering the figure. There is no doubt that Figure 2 is "busier" than Figure A-3, and that as a consequence, Figure 2 takes more effort and time to understand. More generally, the larger the number of quantities of interest an author seeks to present, the busier will be the enhanced bar graph reporting the results. However, the alternative to displaying Figure 2 is to present a graph conveying the general pattern (e.g., Figure A-2) and require readers to look elsewhere for detailed numerical results (e.g., Table 1). We believe this separation of information would impose a substantially larger cognitive load on readers.

Beginning with the next section, we move beyond the specific examples of enhanced bar graphs we have examined to present a set of conventions that underlie the design of the our illustrative enhanced bar graphs and that we believe can guide the design of figures suitable for conveying the results of many political science experiments. **The conventions are printed in bold**; justifications for the conventions—along with illustrations—are in standard text.

Conventions for Displaying \bar{Y} values in an Enhanced Bar Graph

- **The starting point of an enhanced bar graph is a conventional bar graph with a bar for each experimental condition having a length equal to the estimated mean of \bar{Y} in that condition**, as illustrated in Figures 2, A-3 and A-4.¹⁴
- A reader can determine approximate \bar{Y} values by visually matching the length of bars to the tick marks on the axis for the dependent variable. But judging length gets more difficult as bars get farther from the axis. Thus, it is beneficial to readers to **display a numerical point estimate of \bar{Y} within each bar. However, unless a \bar{Y} value is the subject of a hypothesis, do not clutter the figure by displaying a confidence interval for the value.**¹⁵

Conventions for Displaying Estimated Quantities of Interest

The following conventions apply to any quantity of interest in the experiment—typically either a first or second difference in \bar{Y} .

¹⁴ When the dependent variable is binary (and scored 0 or 1), \bar{Y} is equivalent to the proportion of cases for which $Y = 1$ —or when multiplied by 100—to the percentage of cases for which $Y = 1$.

¹⁵ If a researcher has a hypothesis about the value of \bar{Y} in one or more experimental conditions, these values should also be treated as quantities of interest. Accordingly, their text labels should be expanded to include not only a point estimate, but a 95% confidence interval as well (consistent with the conventions in the next section).

- **For each quantity of interest, display a text label containing a point estimate followed by a 95% confidence interval in parentheses** (as in Figures 2, A-3 and A-4). **Denote a first difference in \bar{Y} with the symbol Δ , and a second difference (or “difference between differences”) in \bar{Y} with $\Delta\Delta$.**
- **Mark any point estimate for a quantity of interest that is statistically significant at the 0.05 level with an asterisk**, thereby taking advantage of a familiar convention. Clearly, there is redundancy in reporting a confidence interval and then using an asterisk to indicate statistical significance, since one can judge whether a quantity is statistically significant at the 0.05 level by seeing whether its 95% confidence interval contains zero.¹⁶ However, we believe that an asterisk can facilitate quick recognition by readers about whether a reported quantity is significant without cluttering the figure.
- When the point estimate of a quantity has a trivial magnitude, its precise value is of little importance to a reader. Accordingly, **text labels for estimated quantities that are statistically insignificant or have magnitudes deemed substantively inconsequential should be displayed in smaller and lighter text than labels for quantities that are both statistically and substantively significant.** For example, in Figure 2, point estimates for two first differences are displayed above the bars: $\Delta = 3$ and $\Delta = 9$. The text label for the former—which is viewed as substantively trivial—is smaller and lighter than the label for the latter—which is deemed of practical significance.
- **Include within the text label for each quantity of interest a brief description of its substantive interpretation.** Such interpretations are no substitute for a detailed text discussion of the results. But incorporating brief interpretations into the figure can help readers—especially ones without strong methodological training—understand how estimated treatment effects are derived from the statistical results. For example, in Figure 2, the label for the right-most first difference reported above the bars ($\Delta = 9$) makes explicit to readers that this quantity is an estimate of the “effect of treatment 1 in the presence of treatment 2.” Similarly, the label for the second difference— $\Delta\Delta = 7$, at the center and top of the figure—clarifies that this value reflects a difference in treatment effects.

Conventions for Displaying First Differences in \bar{Y}

- **If the quantities of interest about which there are hypotheses consist of a single first difference in \bar{Y}** (i.e., one is doing a one-factor experiment with a binary treatment), **orient the two bars showing estimated values of \bar{Y} horizontally**—as in Figure A-3.¹⁷ **Then depict the first difference in \bar{Y} with an arrow (i) to the right of the bar for the experimental condition with the smaller value of \bar{Y} , (ii) extending to a vertical dotted line marking the right edge of the other bar, and therefore (iii) with a length equal to the magnitude of the first difference. The direction of the arrow should match the**

¹⁶ This claim is true only when conducting a two-tailed hypothesis test. With a one-tailed test a quantity can be statistically significant at the 0.05 level when its 95% confidence interval contains zero.

¹⁷ Gelman, Pasarica and Dodhia (2002; see also Jacoby and Schneider 2010) argue that *horizontal* bars are more effective than vertical bars at conveying results since the former tend to produce text labels that are easier to read.

direction of change across experimental conditions captured by the first difference.¹⁸ **Display the point estimate for the first difference—using the label “ $\Delta = \#\#$ ”—below the arrow, followed by a 95% confidence interval.**¹⁹ The fact that Figure A-3 (i) makes visually apparent that the length of the arrow is the difference between the lengths of the two bars, and (ii) makes explicit that this difference represents the treatment effect, allows for quick recognition by a reader of the experiment’s main finding.

This approach can be extended to more complex experiments so long as (i) the experimental conditions can be ordered on the vertical axis so that each first difference displayed reflects a difference between the lengths of adjacent bars, and (ii) the quantities of interest do not include a second difference in \bar{Y} . However, when either of these conditions is absent, we propose a different convention for displaying first differences.

- **If the experiment is too complex for the previous convention to be feasible, orient the bars showing estimated values of \bar{Y} vertically—as in Figures 2 and A-4.**²⁰ **Then depict each first difference in \bar{Y} reflecting a treatment effect of interest by connecting the two bars from which the difference is calculated with a U-shaped arrow. The direction of the arrow should match the direction of change across experimental conditions captured by the first difference. If the first difference is statistically significant and with a magnitude believed to be substantively significant as well, the arrow should be solid (as with most of the first differences displayed in Figure 2); if not, the arrow should be dashed, and thus less prominent (as with the upper-left arrow in Figure 2 reflecting the effect of treatment 1 in the absence of treatment 2). Display the point estimate for the first difference—using the label “ $\Delta = \#\#$ ”—inside the arrow, followed by a 95% confidence interval.**²¹ For example, in Figure A-4, consider the first difference in \bar{Y} capturing the effect of level 1 of the treatment:

¹⁸ More formally, the arrow depicting the first difference $\bar{Y}_{\text{condition B}} - \bar{Y}_{\text{condition A}}$ should point from the value of $\bar{Y}_{\text{condition A}}$ to the value of $\bar{Y}_{\text{condition B}}$. Thus, the first difference shown in Figure A-3 is

$$\bar{Y}_{\text{treatment present}} - \bar{Y}_{\text{treatment absent}} = 10.0 - 5.2 = 4.8.$$

If the direction of the arrow in Figure A-3 were reversed, we would instead be depicting the difference,

$$\bar{Y}_{\text{treatment absent}} - \bar{Y}_{\text{treatment present}} = 5.2 - 10.0 = -4.8,$$

which is the additive inverse of the value reported in the figure.

¹⁹ (i) There is debate about whether it is better to derive estimates of first differences with or without regression adjustment (Freedman 2008; Bowers 2011; Lin 2013), but this debate is irrelevant to the best format for displaying these estimates. The estimates displayed in an enhanced bar graph can be predicted values from an econometric model including covariates, or sample statistics not taking into account any covariates.

(ii) Figure A-5 illustrates how Figure A-3 can be modified if the difference in means is too small in magnitude (i.e., the arrow is too short) to permit a text label for the treatment effect to be displayed below the arrow.

²⁰ Given the large set of text labels necessary in an enhanced bar graph with more than two bars, we find that labels are more readable when bars are vertical than when they are horizontal. For a general discussion of factors that should influence the choice between horizontal and vertical bars, see Kosslyn (2006, 53-4).

²¹ (i) We display this information inside the arrow, rather than outside it, to avoid interfering with effective depiction of a second difference (as discussed below).

$$\Delta = \bar{Y}_{\text{treatment level 1}} - \bar{Y}_{\text{no treatment}} = 2.0 - 1.3 \approx 0.8.$$

This difference is depicted by running a U-shaped arrow from the bar for “no treatment” to the bar for “treatment level 1.” **Ideally, all U-shaped arrows depicting first differences should be placed above the bars. However, when this is not feasible without crisscrossing arrows, confusing graphics can be avoided by placing some arrows beneath the bars.** This is also illustrated in Figure A-4, in which there is no way to position an arrow above the bars that runs from the bar for level 1 to the bar for level 3 without cutting across several other arrows. However, by strategically locating the arrows in the figure, we are able to limit the number of arrows that need to be placed beneath the bars to one.²²

- **When an experiment involves comparing a first difference in \bar{Y} across contexts, position the arrows depicting first differences so that any two differences reflecting the same treatment effect in different contexts are depicted with arrows that are aligned horizontally.** This convention allows readers to easily identify treatment effects that can reasonably be compared by scanning horizontally for like-positioned arrows. For example, in Figure 2, we positioned the four first differences displayed so that the two differences measuring the effect of treatment 1 are aligned horizontally, as are the two differences reflecting the effect of treatment 2.

Conventions for Displaying a Second Difference in \bar{Y}

Typically, a second difference (or “difference between differences”) in \bar{Y} is reported to readers only when an experiment is designed to test an interaction hypothesis—that the effect of a treatment varies across contexts. For this reason, when a second difference *is* reported, we believe it should be displayed very prominently within the enhanced bar graph.

- **Depict a second difference in \bar{Y} , $\Delta\Delta$, by connecting the arrows associated with two *first* differences from which the second difference can be calculated with a U-shaped “double arrow.” Display the point estimate for the second difference—using the label “ $\Delta\Delta = \#\#$ ”—inside the arrow, followed by a 95% confidence interval.** This convention is

(ii) In general, we prefer that confidence intervals be portrayed graphically (as in Figure 1) rather than in text format. However, we see no good way of incorporating graphically-depicted confidence intervals for first differences into an enhanced bar graph. And we believe that the benefits of “enhancing” a bar graph far outweigh the benefits of displaying its confidence intervals graphically.

²² Clearly, as the number of treatment effects of interest—and thus the number of arrows and text labels associated with them—grows, the ease with which the information in an enhanced bar graph can be absorbed declines. For example, Figure A-4 takes much more time to understand than Figure A-3. Nevertheless, we believe that Figure A-4 is superior to any other vehicle for conveying the six treatment effects displayed. Note that we do not recommend that authors include arrows and text labels associated with every possible treatment effect that could be calculated simply to make the figure “comprehensive.” For example, if the author of Figure A-4 were interested only in the effect of each treatment level relative to the next lowest level (e.g., level 3 relative to level 2), the number of treatment effects that need to be reported would decline from six to three, and all three could be depicted with arrows above the bars—simplifying the figure.

illustrated at the top of Figure 2. Using a double arrow attracts the reader’s attention, and helps to distinguish the second difference from the first differences denoted by single arrows.²³

Choosing the Order of Bars

The overriding consideration when choosing the order of experimental conditions in an enhanced bar graph is to allow quick and easy pattern recognition.

- **When an experiment examines three or more levels of the same treatment, the order of bars should correspond with the natural ordering of levels, with the lowest level (typically, “no treatment”) at the left** (Kosslyn 2006). This is illustrated—in Figure A-4—for an experiment with four levels of a treatment.
- **When possible, place the two bars associated with a displayed first difference side by side to facilitate visual comparison of the lengths of the bars** (Gelman, Pasarica and Dodhia 2002; Kosslyn 2006). **When it is impossible to accomplish this, and thus some relevant first differences involve nonadjacent bars, place a tick mark within the longer bar at the side closest to the shorter bar to ease visual comparison of the lengths of the two bars.** Figure A-3 reports no first difference involving nonadjacent bars, and so no tick mark is used. However in Figure 2, for example, to help a reader compare the length of the first bar (for both treatments absent) to the length of the third bar (for treatment 1 absent/treatment 2 present), a tick mark denoting the length of the first bar is positioned on the left side of the third bar. Similarly, there is a tick mark in the fourth bar that facilitates its comparison with the second bar.
- **The previous two norms may conflict. When this happens, give priority to preserving the natural ordering of levels.** For example, consider Figure 2. It displays the two first differences reflecting the effect of treatment 2 beneath the bars to avoid a confusing configuration of overlapping arrows and text labels above the bars.²⁴ Each of the first differences beneath the bars involves a comparison of bars that are not side-by-side. We could reduce the number of first differences requiring a comparison of nonadjacent bars from two to one (and fit all arrows above the bars with no crisscrossing required) by reversing the positions of the left-most two bars—as shown in Figure A-6. But much of the value of Figure 2 rests on it making immediately evident by the relative lengths of the bars that

²³ Note that the order of bars in Figure 2 was chosen to facilitate characterizing the second difference as a measure of how the effect of treatment 1 varies depending on whether treatment 2 is present—while also permitting the second difference to be displayed above the bars rather than below. If, instead, we preferred to frame this second difference as a measure of how the effect of treatment 2 varies depending on whether treatment 1 is present—yet still display the second difference above the bars—we would have positioned the two bars for when treatment 1 is absent to the left of the two bars for when treatment 1 is present. Either way, for the benefit of readers without extensive training in research methods, we encourage authors reporting results about interaction between two independent variables in an enhanced bar graph—like that of Figure 2—to include a note in the figure clarifying that the reported second difference could equivalently be calculated in a second way.

²⁴ Placing these two first differences beneath the bars also facilitates the convention (stated above) of aligning horizontally any arrows corresponding to the same treatment effect in different contexts.

although treatment 1 is effective both in the absence, and in the presence, of treatment 2; treatment 1 is more effective when it is provided along with treatment 2. Shifting the order of bars (to produce Figure A-6) would greatly hinder pattern recognition by requiring readers to compare the difference in length of the two bars on the left to the difference in length of the two bars on the right, even though the order of categories within the pairs being compared is switched (Kosslyn 2006).

Conventions for Labeling Axes and Bars

- **Put a text label identifying each experimental condition below its corresponding vertical bar (or to the left of its corresponding horizontal bar, as illustrated in Figure A-3). When doing this would require text that overlaps with the graphics or text labels used to depict one or more first differences, move the text labels identifying experimental conditions to just inside the bars, as in Figures 2 or A-4. Text labels are superior to varying the shading or coloring of bars to identify experimental conditions, which requires a reader to go back and forth between the graph and a legend to match bars with experimental conditions—thereby increasing her cognitive load (Gelman, Pasarica and Dodhia 2002).**
- **The portions of experimental condition text labels that are decisive to distinguishing conditions should be in bold text.** For example, in the labels for experimental conditions in Figure A-3, “absent” and “present” are in bold, but “Treatment”—which is common to both labels—is not.
- **Display at the top of each vertical bar (or at the right end of each horizontal bar) the point estimate for \bar{Y} for the experimental condition. Unless a \bar{Y} value is a quantity of interest for testing a hypothesis, use lighter and smaller text than that used elsewhere to display quantities of interest that are both statistically and substantively significant.**
- **Display at the bottom of each vertical bar (or at the left end of each horizontal bar) “ $n = \#$ ” to report the number of subjects observed for the experimental condition; use small light text so as not to attract a reader’s immediate attention.** Although sample sizes are not the most vital information to be presented, they are relevant for interpreting information about the precision of point estimates, and they can be displayed using very little space and without interfering with the effective communication of the primary quantities of interest.
- **If setting the origin for the Y -axis at zero would leave insufficient space for text labels and arrows, choose a nonzero origin that will make the shortest bar no longer than the minimum required to clearly display the text we suggest be included within each bar (i.e., an estimate of \bar{Y} , a sample size, and perhaps a label for the experimental condition).** We concede that when the origin is substantially greater than zero—as in Figure 2—following this advice risks making relatively small differences between \bar{Y} values appear large. But this risk is lessened by the fact that an enhanced bar graph reports within the graph the specific value of each first difference of interest, making it difficult to misjudge the magnitude of these differences.

Summary of Recommendations about Text Size and Style

To conform to the conventions we introduce above, an author constructing an enhanced bar graph would vary text size based on the importance of the information conveyed.

- **Large text should be reserved for text labels for quantities of interest that are both statistically significant and of a magnitude deemed substantively consequential; such text should also be in bold.**
- **Small text should be used (i) for text labels for quantities of interest that are statistically insignificant or of a magnitude deemed substantively trivial, (ii) for \bar{Y} point estimates, (iii) for sample sizes, (iv) to label each experimental condition, and (v) to label the Y-axis.**

Conclusion

For most experiments conducted by political scientists, the simple conventions described above allow all relevant quantities of interest (both point estimates and confidence intervals) to be conveyed clearly within a single figure. The resulting *enhanced bar graph* combines the pattern recognition strengths of a well-designed graph with the efficient “value look-up” characteristics of a well-structured table.²⁵

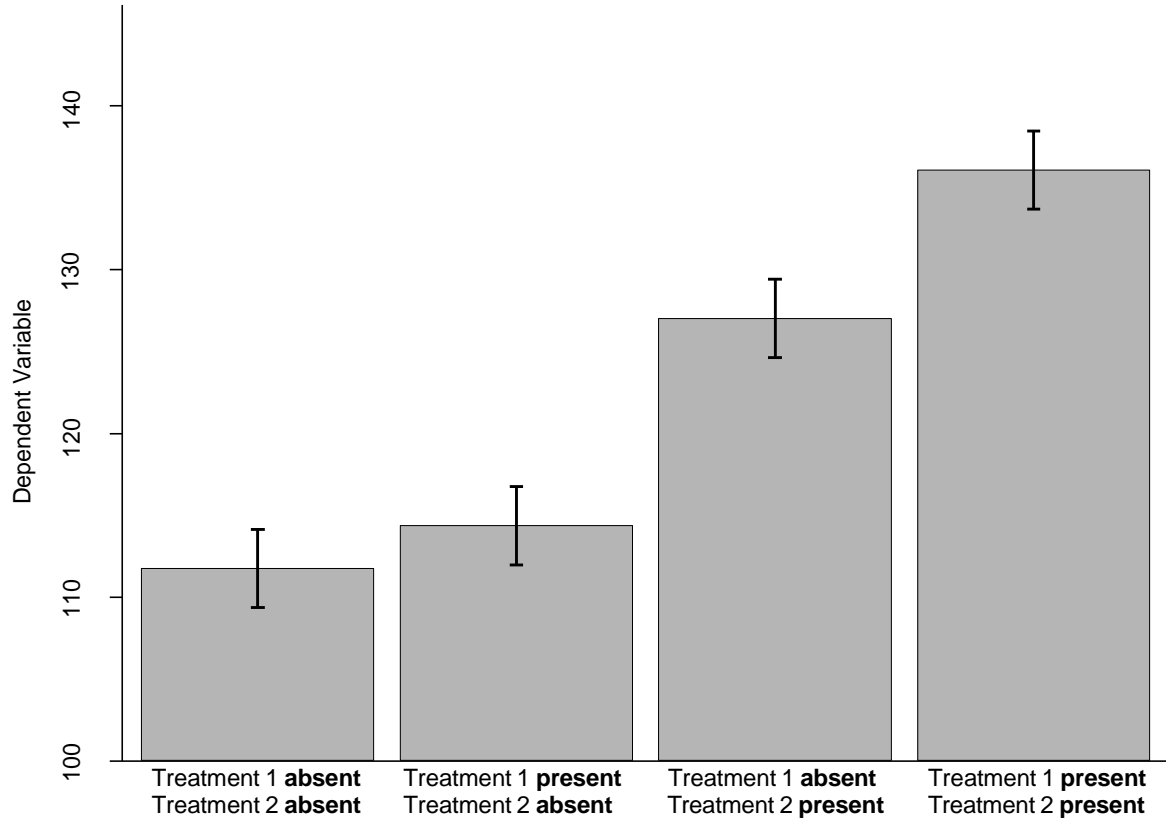
²⁵ Document A-1 examines the strengths and weaknesses of two alternative formats we considered for figures presenting results of experiments—before we ultimately concluded that an enhanced bar graph format is superior.

Table 1. A table showing the results from a fictitious 2 x 2 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) <i>n</i> =20	127.0 (124.6, 129.4) <i>n</i> =20	15.3 (11.9, 18.7)
(B)	Treatment 1 present	114.4 (112.0, 116.7) <i>n</i> =20	136.2 (133.8, 138.5) <i>n</i> =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)

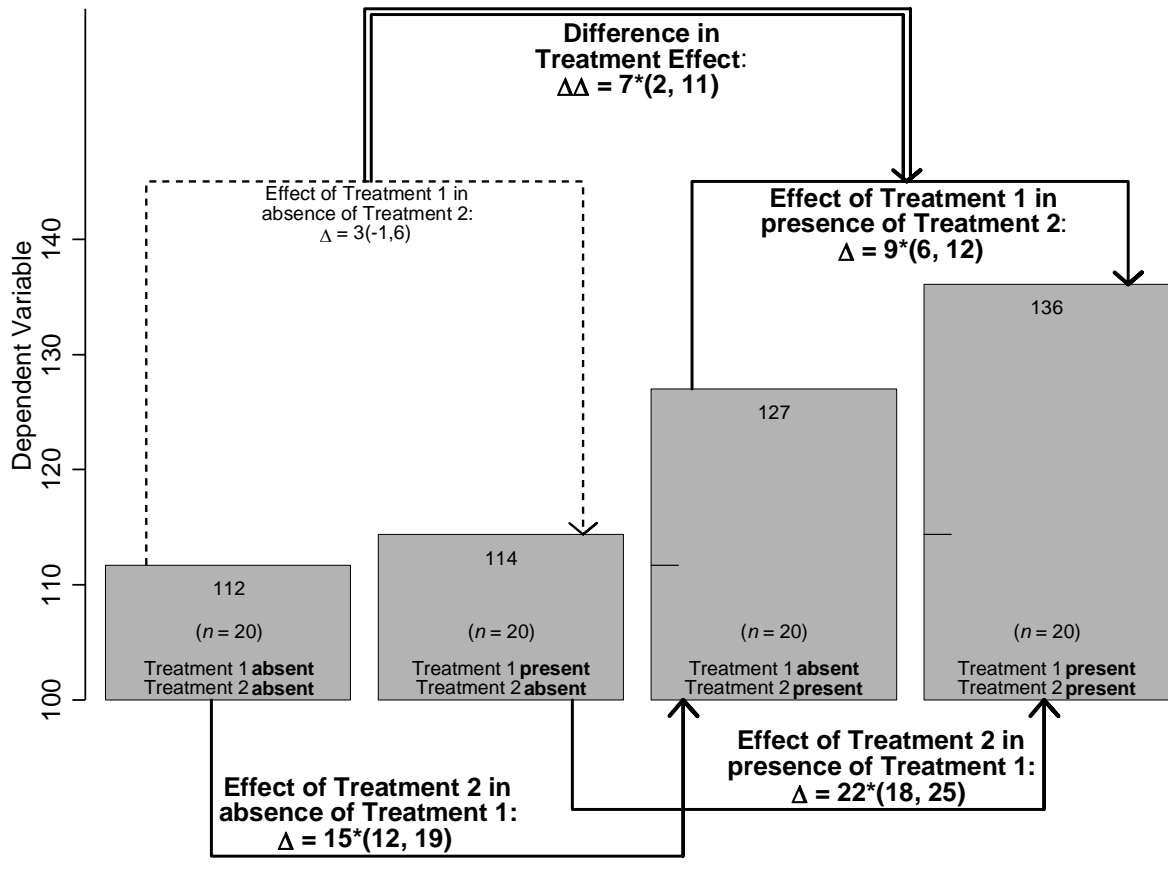
Note: Each of the four non-shaded cells of the table shows the estimated mean of the dependent variable, *Y*, among subjects in an experimental condition. Each row or column marginal (i.e., each lightly-shaded cell) reports a difference in means reflecting the effect of one factor at a value of the other 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.

Figure 1. A conventional bar graph showing the results from Table 1



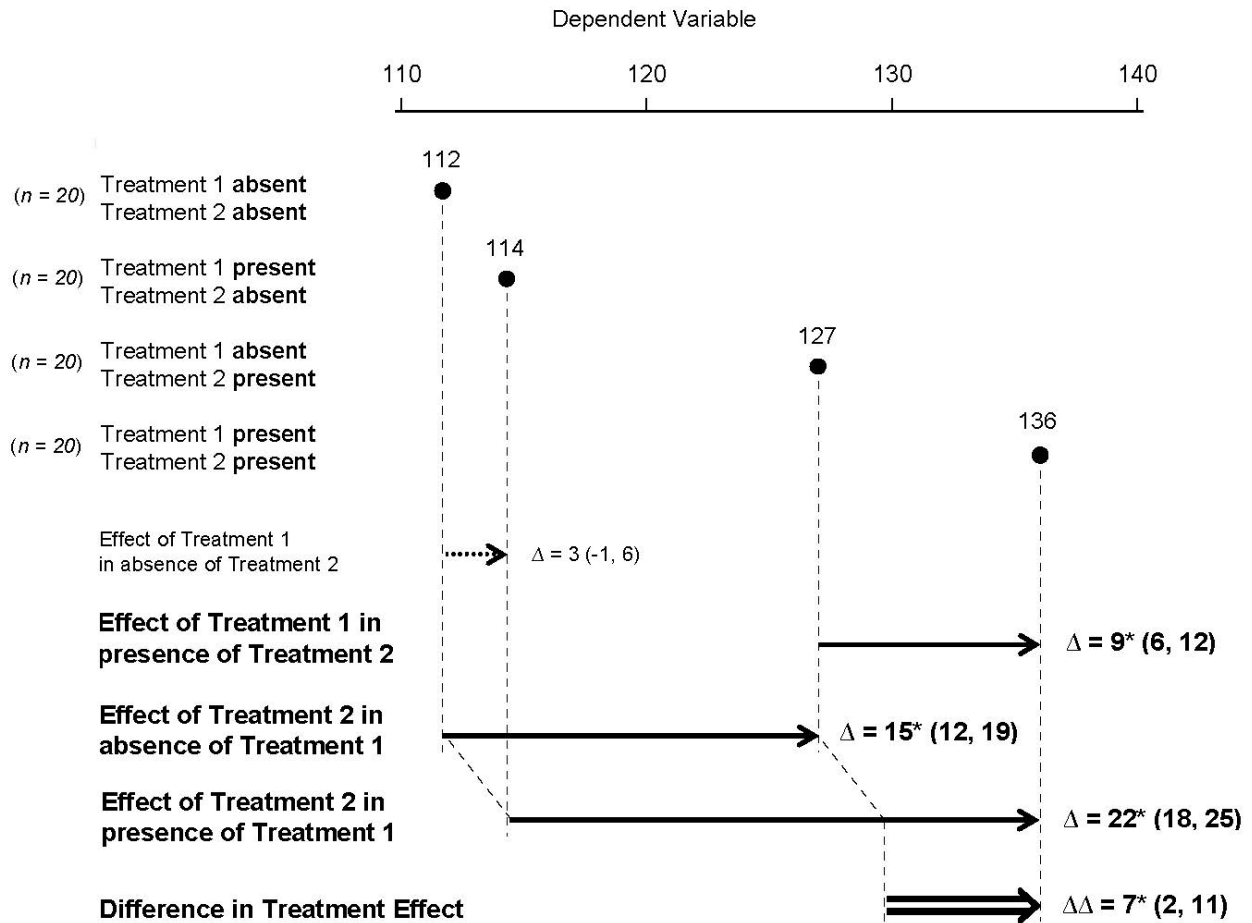
Note: Each bar shows the estimated mean of the dependent variable among subjects in an experimental condition. The vertical line overlaid on a bar shows the boundaries for a 95% confidence interval.

Figure 2. An enhanced bar graph showing the 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, with solid arrows indicating an effect deemed substantively significant. The $\Delta\Delta$ value next to the double arrow is a second difference (i.e., a difference between two differences in means, $7 \approx 9 - 3$) reflecting the strength of interaction. ($\Delta\Delta$ could equivalently be computed 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).

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, with solid arrows indicating an effect deemed substantively significant. The $\Delta\Delta$ 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. ($\Delta\Delta$ could equivalently be computed as a difference between the other two first differences portrayed: $7 \approx 9 - 3$.) 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).

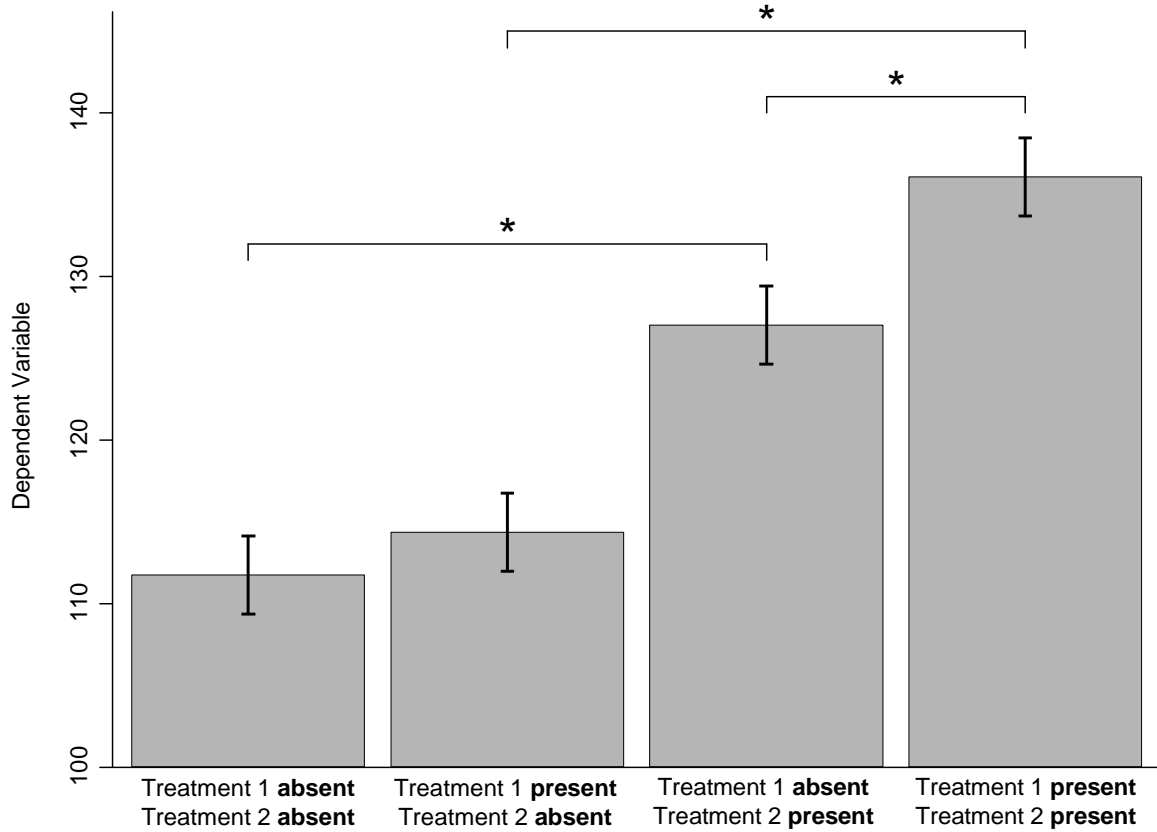
Document A-1

Alternative formats for figures that we considered but ultimately concluded were less effective than an enhanced bar graph

Figure A-7 is an enhanced bar graph showing the same results as Figure 2, but deleting bars depicting the mean of the dependent variable in each experimental condition to permit achievement of the goal of depicting each quantity of interest—the first and second differences in means—with an object whose length equals the quantity. However, we believe this advantage is offset by two consequences of deleting the bars depicting \bar{Y} values. First, the deletion prevents readers from seeing how the value for each first difference is derived from component \bar{Y} values. Second, the deletion makes it harder for a reader to appreciate the practical significance of a first difference value. For example, one gets a far better sense of the magnitude of the effect of treatment 1 in the presence of treatment 2 if one sees not just that the first difference is 9, but also that this value represents the difference between 127 and 136 on Y 's measurement scale.

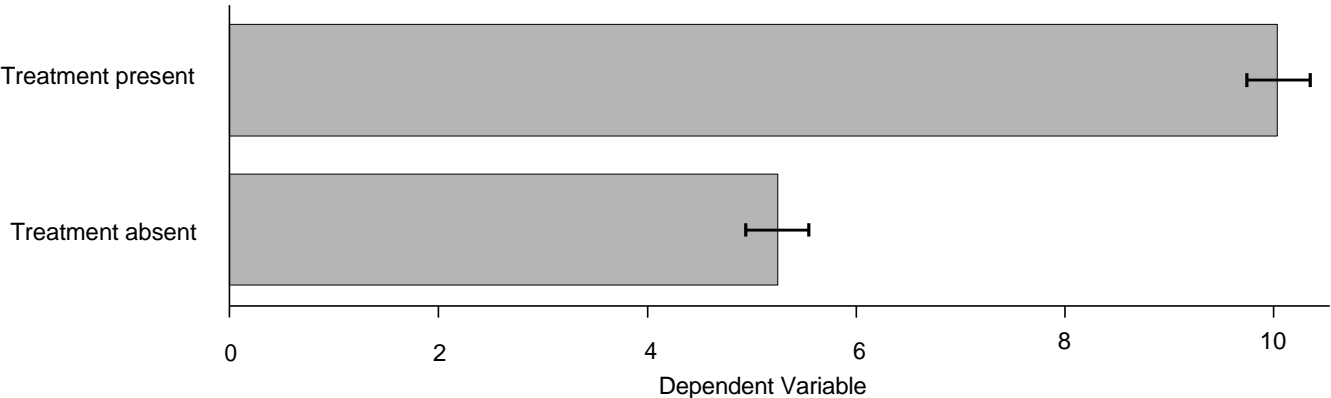
Given the general advantages of dot plots over bar graphs for communicating statistical results (Jacoby 2006; Cleveland 1984), we considered an “enhanced dot plot” taking the form of **Figure A-8** as a format for presenting experimental results. Figure A-8 shows the same results as the enhanced bar graph in Figure 2, adopting all conventions of an enhanced bar graph except that it uses dots instead of bars to depict \bar{Y} values. We concluded that Figure 2 is superior to Figure A-8 since the arrows for depicting first differences in means tend to blend together and lose their visual distinctness when dots are used instead of bars. This weakness in Figure A-8 led us to develop the enhanced dot plot format illustrated in Figure 3, which we believe is better than the format of Figure A-8.

Figure A-1. A modification of the bar graph in Figure 1 to identify statistically significant treatment effects using a bracket/asterisk convention common in some disciplines



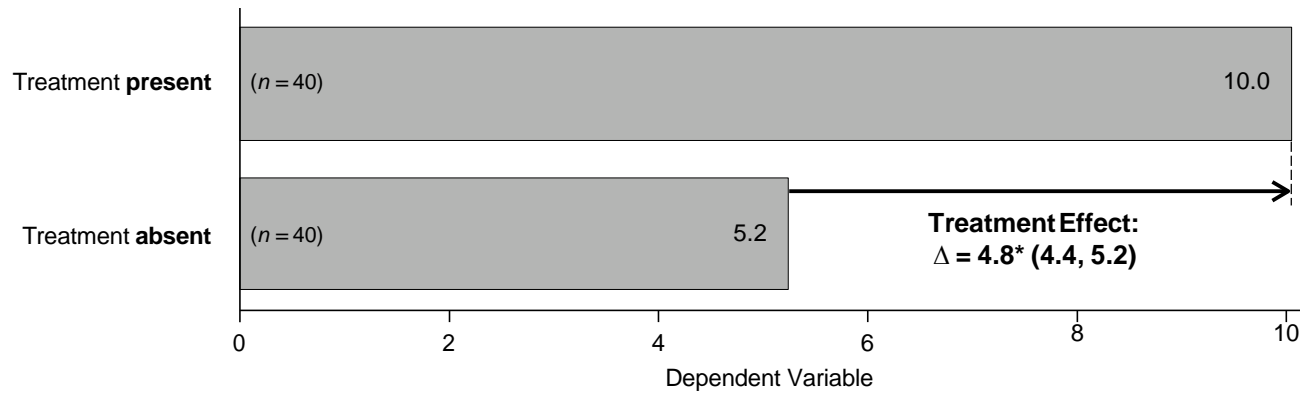
Note: Each bar shows the estimated mean of the dependent variable among subjects in an experimental condition; the vertical line overlaid on a bar shows the boundaries for a 95% confidence interval. Each treatment effect (i.e., first difference in means) that is statistically significant (2-tailed test) at the 0.05 level is denoted by a horizontal bracket extending across the two bars associated with the first difference reflecting the effect, accompanied by an asterisk.

Figure A-2. A conventional bar graph showing the results from a fictitious one-factor experiment with a binary treatment



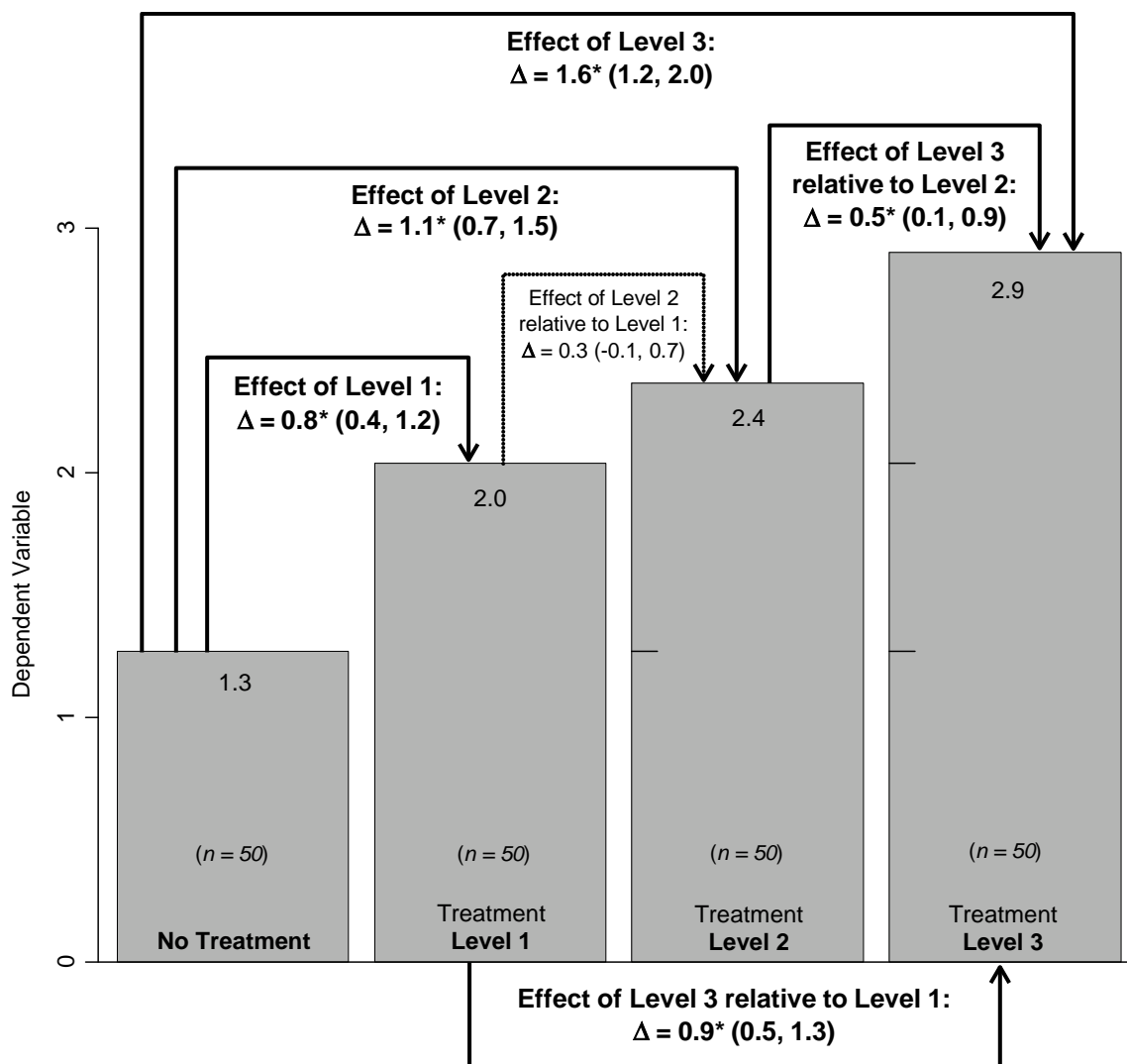
Note: Each bar shows the estimated mean of the dependent variable among subjects in an experimental condition. The horizontal line overlaid on a bar shows the boundaries for a 95% confidence interval.

Figure A-3. An enhanced bar graph showing the results from a fictitious one-factor experiment with a binary treatment



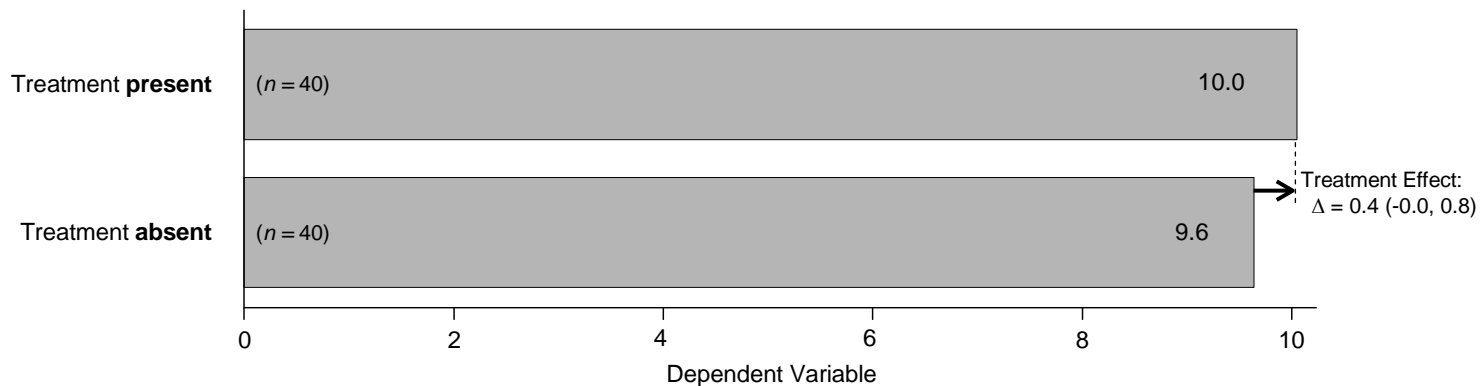
Note: Each bar shows the estimated mean of the dependent variable among subjects in an experimental condition. The Δ value next to the arrow is a point estimate of the difference in means reflecting the effect of the treatment, and is followed by the boundaries for a 95% confidence interval in parentheses. The asterisk (*) indicates statistical significance at the 0.05 level (two-tailed test).

Figure A-4. An enhanced bar graph showing the results from a fictitious one-factor (four-value) experiment to test a hypothesis about the relative effects of alternative levels of a treatment.



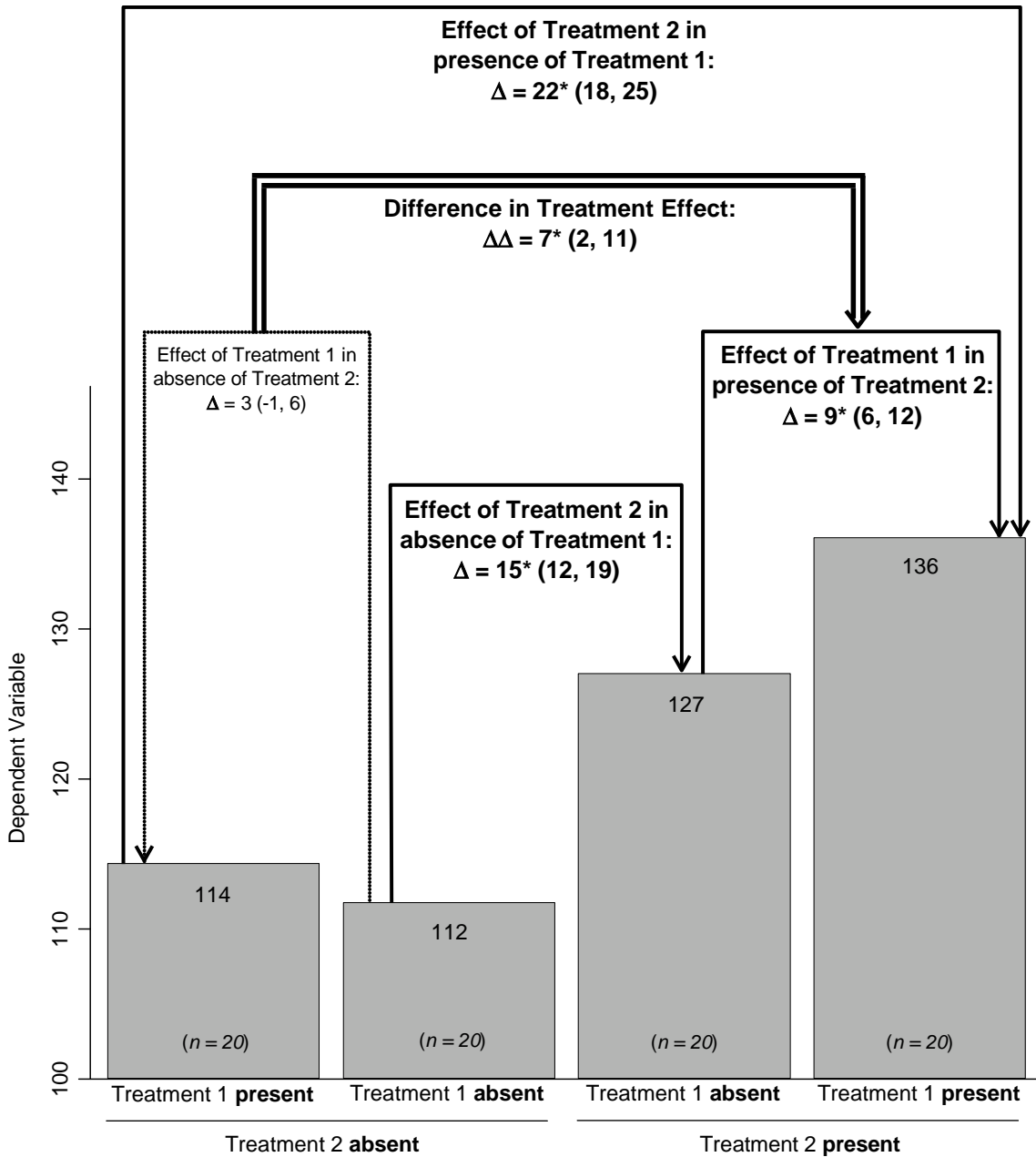
Note: Each bar shows the estimated mean of the dependent variable among subjects in an experimental condition. Each Δ value next to a U-shaped arrow is a first difference in means reflecting the effect of one level of treatment relative to another level (or to the absence of any treatment), with solid arrows indicating an effect deemed substantively significant. 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).

Figure A-5. An enhanced bar graph showing the results from a fictitious one-factor experiment with a binary treatment (when there is not enough room to display the estimated treatment effect in the recommended location)



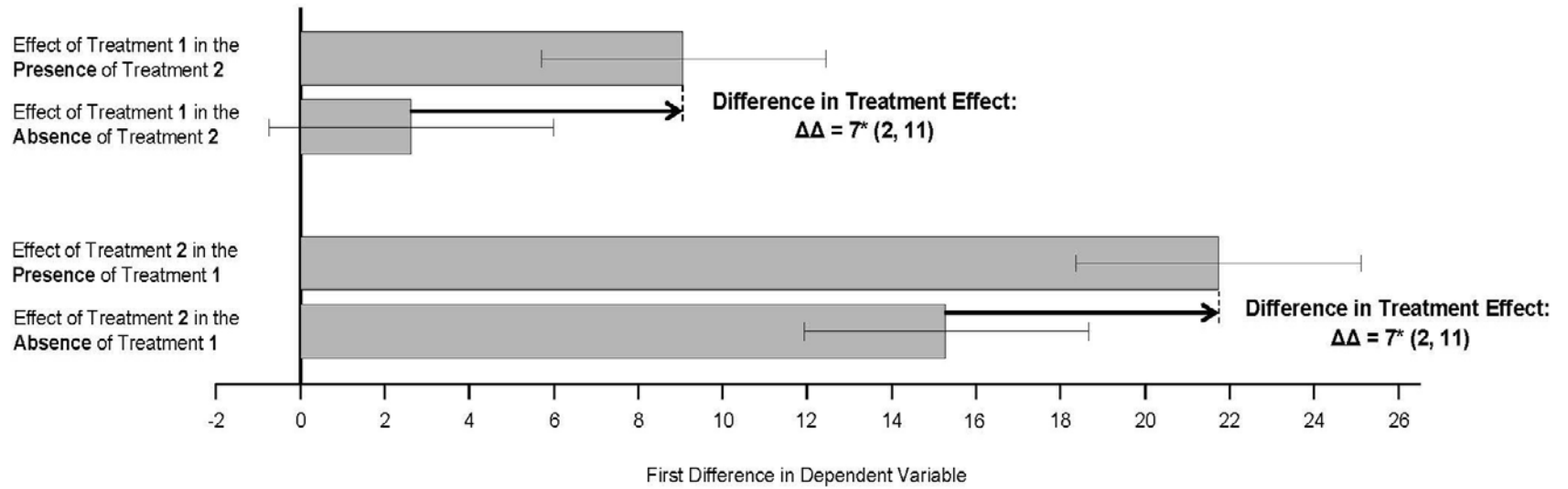
Note: Each bar shows the estimated mean of the dependent variable among subjects in an experimental condition. The Δ value next to the arrow is a point estimate of the difference in means reflecting the effect of the treatment, and is followed by the boundaries for a 95% confidence interval in parentheses.

Figure A-6. An enhanced bar graph showing the same results as Figure 2, but switching the order of bars so as to display as many treatment effects as possible with adjacent bars. (The benefits do not exceed the costs.)



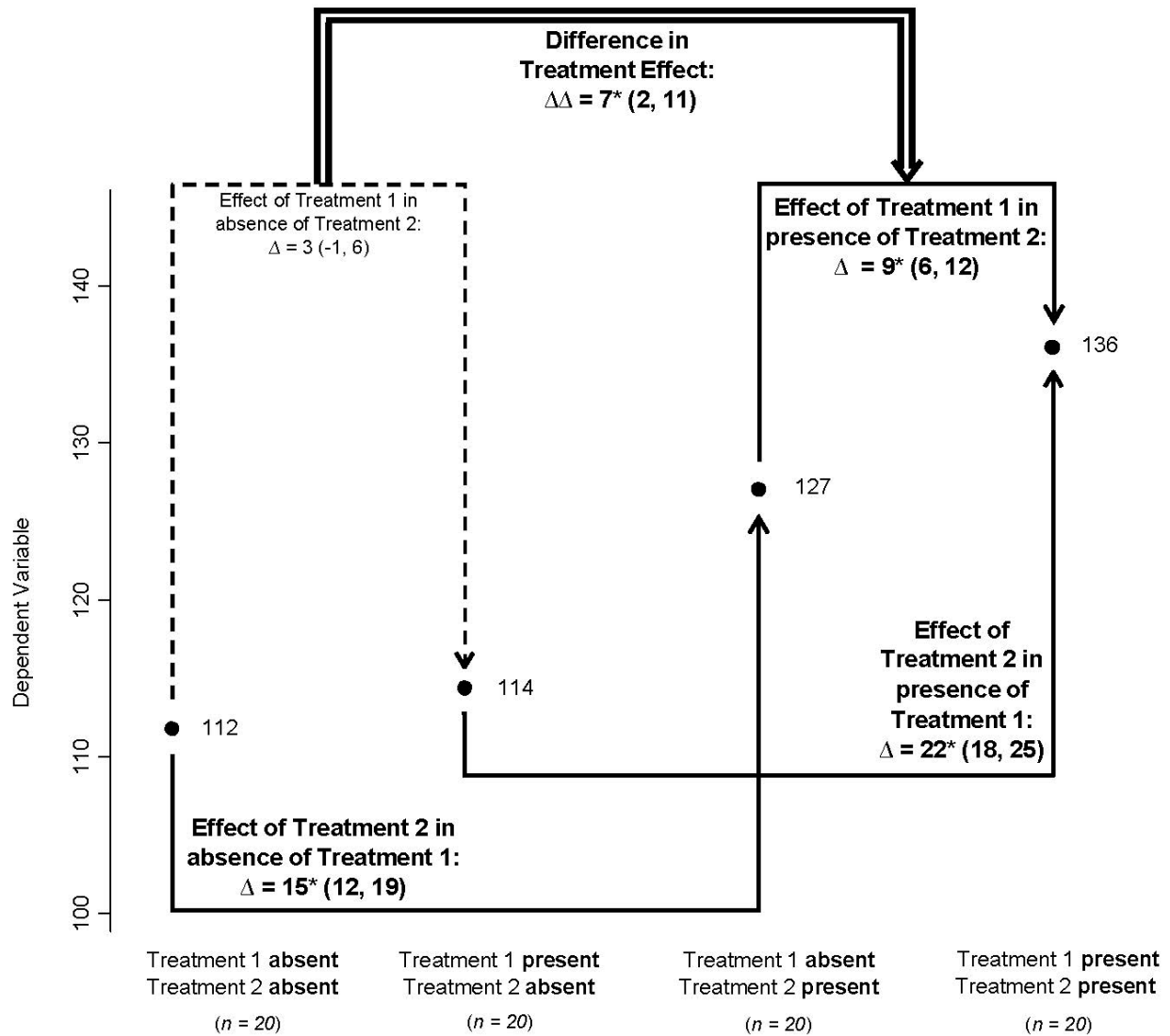
Note: Each bar shows the estimated mean of the dependent variable among subjects in an experimental condition. Each Δ value next to a U-shaped arrow is a first difference in means reflecting a treatment effect, with solid arrows indicating an effect deemed substantively significant. The $\Delta\Delta$ value next to the double arrow is a second difference (i.e., a difference between two differences in means, $7 \approx 9 - 3$) reflecting the strength of interaction. ($\Delta\Delta$ could equivalently be computed 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).

Figure A-7. An enhanced bar graph showing the same results as Figure 2, but deleting bars depicting the mean of the dependent variable in each experimental condition to allow for better graphical representation of a second difference as a difference between two first differences.
(The benefits do not exceed the costs.)



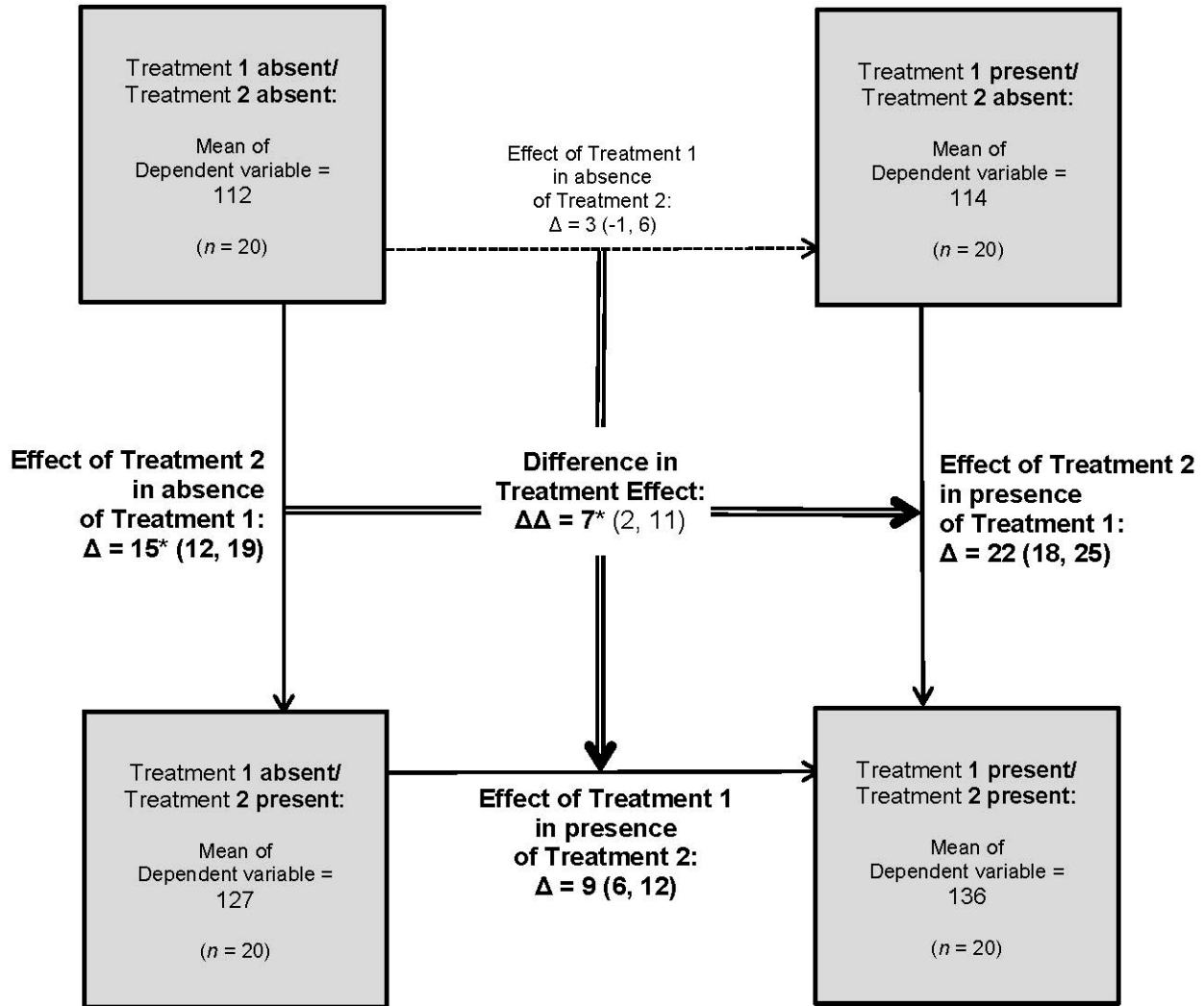
Note: Each bar shows a first difference—i.e., the difference between the estimated mean of the dependent variable across two experimental conditions—reflecting a treatment effect. The horizontal line overlaid on each bar shows the boundaries for a 95% confidence interval. The $\Delta\Delta$ value next to each arrow is the second difference (i.e., a difference between two differences in means) reflecting the strength of interaction, and is followed by the boundaries for a 95% confidence interval in parentheses. An asterisk (*) indicates statistical significance at the 0.05 level (two-tailed test).

Figure A-8. An “enhanced dot plot” showing the same results as Figure 2 using U-shaped 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 next to a U-shaped arrow is a first difference in means reflecting a treatment effect, with solid arrows indicating an effect deemed substantively significant. The $\Delta\Delta$ value next to the double arrow is a second difference (i.e., a difference between two differences in means, $7 \approx 9 - 3$) reflecting the strength of interaction. ($\Delta\Delta$ could equivalently be computed 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).

Figure A-9. An alternative figure using text boxes (and no graph elements) to display the same results as Figure 2



Note: Each shaded square shows the estimated mean of the dependent variable among subjects in an experimental condition. Each Δ value next to an arrow is a first difference in means reflecting a treatment effect, with solid arrows indicating an effect deemed substantively significant. The $\Delta\Delta$ value in the center of the figure is a second difference (i.e., the difference between two differences in means, $7 = 22 - 15 \approx 9 - 3$) reflecting the strength of interaction. 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).

Table A-1. Characteristics of the eleven articles in 2014 issues of *Journal of Experimental Political Science (JEPS)* that present experimental results and report at least one treatment effect

Issue	Author(s)	Text Only? ¹	Display Format ²	Notes
1	Al-Ubaydli, McCabe and Twieg	No	Graph	see Figure 4 (51)
1	Bolsen, Druckman and Cook	No	Table	see Table 1 (30-31)
1	Brancati	Yes (9)	--	
1	Krupnikov and Levine ³	No	Table	see Table 3 (68)
1	Stadelmann, Portmann and Eichenberger	No	Both	see Figure 1 (19) & Table 1 (20)
2	Broockman	No	Table	see Table 3 (112)
2	Brown	No	Table	see Table 4 (141)
2	Coppock	No	Graph	see Figure 1(a) (166)
2	Healy, Kuo and Malhotra	No	Both	see Figure 1 (155) & Table 4 (154)
2	Mironova and Whitt	No	Table	see Table 1 (176)
2	Woon	No	Table	see Table 1 (192)
	COLUMN DISTRIBUTION:	Yes: 1 No: 10	Graphs only: 2 Tables only: 6 Both: 2	

Notes: An article is deemed to report one or more treatment effects if it either reports a point estimate of the effect or indicates whether the effect is statistically significant. Numbers in parentheses in any column indicate a page number from the *JEPS* article at which the indicated information is found.

- (1) Coded as “Yes” only if an article discusses treatment effects in the text, but no treatment effect is displayed in a graph or table.
- (2) “Both” indicates that at least one treatment effect is displayed in both a graph and a table. “Graph” indicates that at least one treatment effect is displayed in a graph and no treatment effects are displayed in a table. “Table” indicates that at least one treatment effect is displayed in a table and no treatment effects are displayed in a graph. (Tables of raw regression coefficients are not coded as presenting treatment effects in tables.)
- (3) Krupnikov and Levine (2014) present four experiments. This table describes Experiment 1. However, the coding would be identical for each of the other experiments.

References

- Aidala, Zachary, Rebecca Croston, Jessica Schwartz, Lainga Tong and Mark E. Hauber. 2015. "The Role of Egg–nest Contrast in the Rejection of Brood Parasitic Eggs." *Journal of Experimental Biology* 218: 1126-36.
- 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.
- Arceneaux, Kevin and Martin Johnson. 2015. "How Does Media Choice Affect Hostile Media Perceptions? Evidence from Participant Preference Experiments." *Journal of Experimental Political Science* 2:12-25.
- Bowers, Jake. 2011. "Making Effects Manifest in Randomized Experiments." In *Cambridge Handbook of Experimental Political Science*, eds. Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia. New York, NY: Cambridge University Press.
- 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.
- Chaterjee, Kaushik, Steve Hung, Girish Kumar and Carl G. Simon Jr. 2012. "Time-Dependent Effects of Pre-Aging 3D Polymer Scaffolds in Cell Culture Medium on Cell Proliferation." *Journal of Functional Biomaterials* 3:372-81.
- Cleveland, William S. 1984. "Graphical Methods for Data Presentation: Full Scale Breaks, Dot Charts, and Multibased Logging." *The American Statistician* 38:270-80.
- Cole, T.J. 2015. "Too Many Digits: The Presentation of Numerical Data." *Archives of Disease in Childhood* 100:608-9.
- Darby, Kevin P. and Vladimir M. Sloutsky. 2015. "The Cost of Learning: Interference Effects in Memory Development." *Journal of Experimental Psychology* 144:410-31.
- DiPaola, Christian P., Matthew J. DiPaola, Bryan P. Conrad, Mary Beth Horodyski, Gianluca Del Rossi, Andrew Sawers, and Glenn R. Rechtine II. 2008. "Comparison of Thoracolumbar Motion Produced by Manual and Jackson-table-turning Methods." *Journal of Bone and Joint Surgery* 90:1698-1704.
- 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.
- Freedman, David A. 2008. "On Regression Adjustments to Experimental Data." *Advances in Applied Mathematics* 40:180–93.
- Gelman, Andrew, Cristian Pasarica and Rahul Dodhia. 2002. "Let's Practice What We Preach: Turning Tables into Graphs." *The American Statistician* 56:121-30.
- Gelman, Andrew and Hal Stern. 2006. "The Difference Between 'Significant' and 'Not Significant' is not Itself Statistically Significant." *The American Statistician* 60:328-31.
- 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.
- Hirashima, Toshiko, Wook-Cheol Kim, Kouei Kawamoto, Takashi Yoshida and Toshikazu Kubo. 2009. "Evaluating Bone Union of Distal Radius Fractures by Measuring Impedance Values." *Orthopedics* 32 DOI: 10.3928/01477447-20090101-10.
- Jacoby, William G. 2006. "The Dot Plot: A Graphical Display for Labeled Quantitative Values." *The Political Methodologist* 14:6-14.
- Jacoby, William G. and Sandra Schneider. 2010. "Graphical Displays for Political Science Journal Articles." Paper presented at the Visions in Methodology Conference, Iowa City, March 2010.
- 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.
- Krantz, David H. 1999. "The Null Hypothesis Testing Controversy in Psychology." *Journal of the American Statistical Association* 94:1372-81.
- 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.
- Lin, Winston. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *Annals of Applied Statistics* 7:295-318.
- 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.
- Tufte, Edward R. 1983. *The Visual Display of Quantitative Information*, 2nd ed. Cheshire, CT: Graphics Press
- Wainer, Howard. 1997. "Improving Tabular Displays, with NAEP Tables as Examples and Inspirations." *Journal of Educational and Behavioral Statistics* 22:1-30.