AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

2022, Vol. 28, No. 2, 314–328
https://doi.org/10.1037/xap0000393

Graphs Do Not Lead People to Infer Causation From Correlation

Madison Fansher, Tyler J. Adkins, and Priti Shah
Department of Psychology, University of Michigan

Media articles often communicate the latest scientific findings, and readers must evaluate the evidence and consider its potential implications. Prior work has found that the inclusion of graphs makes messages about scientific data more persuasive (Tal & Wansink, 2016). One explanation for this finding is that such visualizations evoke the notion of "science"; however, results are mixed. In the current investigation we extend this work by examining whether graphs lead people to erroneously infer causation from correlational data. In two experiments we gave participants realistic online news articles in which they were asked to evaluate the research and apply the work's findings to a real-life hypothetical scenario. Participants were assigned to read the text of the article alone or with an accompanying line or bar graph. We found no evidence that the presence of graphs affected participants' evaluations of correlational data as causal. Given that these findings were unexpected, we attempted to directly replicate a well-cited article making the claim that graphs are persuasive (Tal & Wansink, 2016), but we were unsuccessful. Overall, our results suggest that the mere presence of graphs does not necessarily increase the likelihood that one infers incorrect causal claims.

Public Significance Statement

Prior work suggests that merely showing people a graph impacts their interpretation of an accompanying message because graphs are associated with science. The present work however suggests that graphs do *not* significantly influence people's evaluations. Understanding how graphs influence reasoning about scientific data is important given than the media often uses graphs to explain scientific data to the public.

Keywords: scientific reasoning, persuasion by graphs, data visualization

Supplemental materials: https://doi.org/10.1037/xap0000393.supp

An important component of scientific reasoning is understanding the difference between correlation and causation (Bleske-Rechek et al., 2015; Norris et al., 2003; Shah et al., 2017). Failing to recognize the difference between correlation and causation may lead to ill-informed decision-making with minor to catastrophic consequences. For example, there may be a temporal correlation between a diagnosis of autism and the proximity of childhood vaccinations. However, one should be able to reason that childhood vaccinations and signs of autism often occur around the same age, and that the third variable of age explains the observed relationship as there is no scientific evidence that vaccines cause autism (DeStefano & Shimabukuro, 2019). Unfortunately, researchers have found that people frequently mistakenly assume there to be causal relationships based on correlational data when this conclusion is unwarranted (Bleske-Rechek et al., 2015; Klaczynski et al., 1997; Norris et al., 2003; Rodriguez et al., 2016) and fail to consider

This article was published Online First February 28, 2022.
Madison Fansher https://orcid.org/0000-0003-0009-4337
Tyler J. Adkins https://orcid.org/0000-0003-1472-167X
Priti Shah https://orcid.org/0000-0001-5908-0453
The authors would like to thank Audrey Michal for contributing to study design and the Institute of Educational Sciences (IES) for funding this research (Award # R305A170489).

Supplemental materials including survey questions and article stimuli, the data for Experiments 1–3, and the analysis code are available through the Open Science Framework (OSF) at https://tinyurl.com/graphs2021. Experiments 1–3 were preregistered, and we acknowledge that there was slight deviation from the preregistered analysis. We originally preregistered that we would conduct all our analyses in Experiments 1 and 2 twice, once with attitude toward science (ATS) as the belief in science moderator and once with Tal and Wansink's (2016) single item as the belief in science moderator. In Experiment 2 we forgot to include the single item belief in science measure, so we decided to only run the models with ATS as the predictor. In Experiment 2, we deviated from the data collection method in that we hoped to collect data from participants until we were satisfied with the uncertainty around the model estimates. In Experiment 1 we saw little improvement in the size of our models' 95% highest posterior density intervals after doubling sample size, thus, we saw little value in doubling our sample size again for Experiment 2, as we did previously in Experiment 1. Lastly, in the preregistration for Experiment 3 we originally intended to analyze the data with traditional null-hypothesis testing methods. Instead, we opted for Bayesian methods as they allow us to make inferences about null effects. For all three experiments we did not originally intend to use Bayes factors, thus we preregistered that we would use uninformative flat priors across all analyses. It was imperative that we instead use more informative priors as Bayes Factors are undefined for models with flat priors (Gelman et al., 2013).

Correspondence concerning this article should be addressed to Madison Fansher, Department of Psychology, University of Michigan, 530 Church Street, Rm 1004, Ann Arbor, MI 48109, United states. Email: mfansher@umich.edu

the role of third variables when evaluating such data (Klaczynski et al., 1997; Shah et al., 2017). Causality bias is often exacerbated by media reports that speculate about causality (Barrowman, 2014). As such, it is important to consider the factors that may influence everyday scientific reasoning about correlational data. Here, we consider one factor that may influence whether one interprets or is persuaded to believe correlational data as causal, that is, the presence of graphs.

Influence of Graphs on Scientific Reasoning

Some studies find that depicting data graphically makes quantitative information more salient or influential in their decision-making. For example, Fagerlin et al. (2011) found that when participants were presented with narrative testimonials and data on the side effects of a medical treatment, they were more likely to rely on the data instead of the anecdotes when they were presented with the information graphically. In a similar vein, Chua et al. (2006) found that people were willing to pay more money to reduce risk when the risks were represented with graphs compared to numbers. This result held even when both foreground (number of risky outcomes) and background (full sample/denominator) were presented, suggesting that the effect of the presence of graphs was not merely due to reducing attention to the denominator.

Other research finds that the presence of graphs influences scientific reasoning. Tal (2015) argues that the presence of a graph gives claims a "scientific halo" which makes the presented claim more convincing. One study that received a great deal of media attention was reported by Tal and Wansink (2016). In this study, participants read a short vignette about the effectiveness of a medication and were assigned to view the vignette either alone or accompanying a bar graph of the data. Participants were then asked to rate the effectiveness of the medication and to answer the question: "Does the medication really reduce illness?." They found that participants who saw a bar graph rated the medication as more effective than participants who did not see the bar graph, and that more people in the bar graph condition believed that the medication would truly reduce illness (96.55% vs. 67.74%). In another study, Heyer et al. (2020) had participants reason about data driven narratives on the U.S. opioid epidemic. They found that viewing data graphically led to greater attitude change compared to text-only, illustrating that graphs can be moderately influential even for controversial topics.

Although these data provide evidence that graphs influence scientific reasoning in some contexts, other work suggests the opposite or suggests that the effect of data visualizations only emerges under specific circumstances. Dragicevic and Jansen (2018) attempted a direct replication of Tal and Wansink (2016) and were unable to replicate the original finding that graphs increased the perceived effectiveness of a medication. McCabe and Castel (2008) investigated how different types of visualizations, such as bar graphs, influenced the perceived credibility of cognitive neuroscience research. They gave participants fictional news articles with either no image, a brain image, or a bar graph depicting the data. Participants were then asked to rate how well the article was written, whether the title of the article was a good description of the results, and whether the scientific reasoning in the article made sense. They consistently found that participants shown brain images rated the scientific reasoning of the article to be higher than the other two groups, but they did not find a difference between text-only and bar-graph conditions on the perceived scientific reasoning in the article. Pandey et al. (2014) found that graphical depictions of data have the greatest persuasive effect when the viewer has a neutral attitude toward a topic, and Tal and Wansink (2016) found in a follow-up experiment that the presence of a bar graph was most influential when participants scored high on an item measuring belief in science.

Mechanisms in Which Graphs Influence Cognition

Although the literature on the impact of graphs on reasoning about data is divided, there are multiple proposed cognitive mechanisms by which visualizations might influence evidence evaluation. Some researchers argue that graphs are inherently "scientific" and act as an indicator of science (Tal & Wansink, 2016). Haard et al. (2004) suggest that if something is scientific, this leads to the assumption of sophistication, the inference of expertise, and therefore the assumption of source credibility. More specifically, Tal and Wansink (2016) described the effect of graphs as the following inferential process: "The information contains a graph (premise); Graphs signal a scientific basis (premise); Therefore, the information has a scientific basis (conclusion); A scientific basis indicates truth (premise); Therefore, the information is true (conclusion)" (p. 123). Similarly, McCabe and Castel (2008) suggest that the inclusion of visualizations (graphs and brain images) allows the reader to infer more scientific value from the text, as people tend to associate graphical depictions of data with hard sciences (McCabe & Castel, 2008; Smith et al., 2002). A related proposed mechanism is that people in general seek and are more satisfied with "low-level" or "reductionist" explanations which simple visualizations may provide (Rhodes et al., 2014; Weisberg et al., 2008).

Another set of proposed mechanisms focuses on the idea that visualizations may be engaging or aesthetically pleasing. Lauer and O'Brien (2020) suggest that the color, contrast, and other features of graphs may make them especially compelling, leading them to be the focal point leaving text to fade in the background. As such, graphs may also act as "seductive details" which are related but unessential peripheral details that make text more appealing to the reader (Harp & Mayer, 1998). Lipkus and Hollands (1999) suggest, similarly, that graphs may attract and hold attention because the information is presented as a concrete display.

Alternatively, graphs may influence cognition by aiding with comprehension of quantitative information (see Hegarty, 2011). Supporting this idea, van der Linden et al. (2014) found that presenting participants with descriptive text and simple data visualizations (i.e., pie charts) led to better understanding of the scientific consensus on climate change, with the presence of data visualizations being especially beneficial for Republicans. Although Dragicevic and Jansen (2018) were unable to replicate Tal and Wansink (2016)'s finding that there was a persuasive effect of graph presence, they did find that the presence of a graph increased perceived understanding of a drug's efficacy. Graphs may improve understanding by revealing data patterns that may go unnoticed otherwise and thus may assist with viewing and interpreting quantitative information (Lipkus & Hollands, 1999). In general, the mechanism by which graphs increase understanding is that they make some comparisons salient by organizing information (Franconeri et al., in press) and reducing cognitive load (Card et al., 1999; Larkin & Simon, 1987). While there are multiple proposed mechanisms by which graphs might influence scientific reasoning, these mechanisms are not all mutually exclusive. For example, if a visualization is aesthetically pleasing, a reader may spend more time engaging with the content, leading to increased understanding of the data.

Graphs and Reasoning About Correlational Data

Given these findings, we wondered whether the inclusion of graphs increases the likelihood that one infers causality from correlational data. If graphs do affect evaluations of data, as reviewed in the literature above, viewing a graph of correlational data may lead the reader to believe that the relationship presented is valid and may influence the perceived strength of the relationship between the presented variables. If graphs act as indicators of science, it is possible that including such visualizations increases the perceived credibility of the research and its conclusions, which may increase the likelihood that one erroneously infers causation from correlational data, as a causal mechanism is the "strongest" relationship between two variables. In alignment with this hypothesis, Xiong et al. (2020) studied how visual characteristics such as graph type and data aggregation level influence whether people interpret the relationship between two variables as correlational or causal. They compared ratings of perceived correlation and causality between different types of visualizations including bar graphs, line graphs, and scatter plots. The visualizations shown to participants systematically varied in the amount of data aggregation, that is, the number of points used to illustrate the same data set. Across a series of three experiments participants were more likely to infer causality from scatter and line graphs in comparison to bar graphs. The researchers suggest that line and scatter plots may have led to a greater assumption of causality because they are associated with continuous data trends. They also found that the greater the data aggregation (i.e., data binned into fewer bars/points), the more likely it was that participants inferred a causal relationship. However, it is also worth noting that the presence of a graph did not lead to higher ratings of correlation nor causation when compared to a text-only condition.

The Present Study

Although the prior literature on the influence of graphs on making inferences about data is mixed, given Tal and Wansink's (2016) finding on the influence of graphs and Xiong et al.'s (2020) examination of the specific conditions under which people infer a causal relationship when viewing data visualizations, the present study investigates whether the effect of such visualizations emerges when readers reason about correlational data. Much of the prior work on the impact of graphs on interpretations or evaluations of data has been in the context of fictional decision-making scenarios (e.g., how much would you pay for this brand of toothpaste, or which medical treatment would you select) or judgments about magnitudes (e.g., how much better is drug A compared to drug B). The present study expands upon this work by studying the impact of the presence of graphs in the context of everyday scientific reasoning. Participants read about correlational studies publicized by a fictional online news outlet, "Scientific Citizen" in which the content of the article was formatted to be consistent with a news article one would come across in their everyday browsing. Furthermore, we expand on prior work by having participants apply the conclusions of the articles to hypothetical scenarios in which the article's results have implications for improving a person's depression (Experiment 1) or improving a person's physical fitness (Experiment 2).

In two studies, participants read about a correlational study and were randomly assigned to view the text only or text with a corresponding graph. We examined how the presence of graphs influenced reasoning about the perceived strength, causality, and scientific basis of the relationship between two variables, as well as whether participants would apply the article's conclusion to a real-life scenario. To test whether graphs are influential *because* they act as indicators of science (Tal & Wansink, 2016), we also measured attitude toward science (ATS) and assessed whether ATS interacted with the presence of a graph for all outcome variables. Given prior work, we hypothesized that graphs would influence participants to perceive causality from correlational data. We also hypothesized that there would be an interaction between ATS and graph presence for all our dependent variables such that people with high ATS would be more influenced by the presence of a graph.

In Experiment 1, the fake data were situated in the context of a correlational study showing a relationship between attending religious service and happiness; in Experiment 2, the fake data were situated in a more neutral scenario featuring a relationship between attending sporting events and physical fitness. Lastly, in Experiment 3, we attempt a direct replication of Tal and Wansink (2016). Across the three experiments, we find little evidence for an impact of graphs when participants reasoned about science in both novel (correlational) and previously studied (experimental) contexts.

Experiment 1

The goal of Experiment 1 was to examine whether the presence of a graph would influence reasoning about correlational data. Participants read a short vignette about a correlational study suggesting a positive relationship between attending religious service and happiness. This topic was chosen as it was a realistic news headline for a correlational study. At the time of the study's conception, a study from Pew Research Center (2019) showing that people who attend church tend to be happier was widely publicized in online media outlets. Interestingly, these online news sources either explicitly discussed the fact that the study did not show a causal link between attending church and happiness (Kuruvilla, 2019), or withheld this critical piece of information (Klett, 2019). We assigned participants to different visualization conditions and had them reason about the presented data in a made-up online news article. We hypothesized that there would be an influential effect of graphs along with interactions between graph presence and ATS for all outcome variables such that people high in ATS would be more influenced by the presence of a graph.

Method

Participants

A group of 803 participants were recruited from Amazon Mechanical Turk to participate in an online study, Age, *M* (*SD*): 34.89 (10.55) years; 33.34% Female, 66.5% Male, .001% Other. Workers had to be located in the U.S. to participate in our study and they were compensated $1 for their participation. All procedures were determined to be exempt by the University of Michigan Institutional Review Board and preregistration for Experiment 1 may be viewed at https://aspredicted.org/blind.php?x=jn3xa3

Design

The experiment employed a between-subjects design where participants were randomly assigned to one of four groups: a text-only condition, text plus a 2-bar bar graph (bar graph condition), text plus a line graph with data aggregated into 2 points (2 point line graph condition), or text plus a line graph with data aggregated into 6 points (6 point line graph condition). The types of graphs included in the various conditions were based on work by Xiong et al. (2020) as they found that line graphs with data aggregated into 2 or 6 points led to high ratings of perceived causality between two variables. While scatter plots may also be used to illustrate correlational data, Xiong et al. (2020) found no difference between line and scatter plots on perceived causality, thus we only include line graph stimuli for the sake of simplicity. We also include a 2-bar bar graph condition because this is the type of graph displayed to participants in Tal and Wansink (2016).

Materials

All materials were presented to participants via Qualtrics survey software. The main stimuli were realistic online news articles created by the experimenter for the made-up source "Scientific Citizen." The news articles described a study showing a link between attending religious service and happiness. Participants read the following vignette:

> Researchers at University of Michigan surveyed 400 senior students on their daily habits and mental health. The project was part of the University's initiative to improve the mental health services offered to undergraduate students. The researchers found that 65% of the respondents attended religious service at least once in the last year. When students were asked to rate their happiness on a 7-point scale, there was a positive correlation of .96 between reported happiness and the number of times students reported attending religious service in the last year. Participants who did not attend religious service had an average happiness rating of 2.5, while participants attending religious service 50 or more times had an average rating of 6. This research suggests that attending religious service can lead to increased happiness and that religious service should be explored as a treatment option for those with mood disorders.

It is important to note that participants in the graph conditions had access to more information about the data than those in the text-only condition, as a large amount of information may be depicted via visualization that is difficult or impossible to depict with only text. Even in the bar graph condition where only two data points were displayed, participants were able to visually examine the size of the effect. To try to better equate the amount of information participants had about the data across conditions, we made sure to highlight key information about the data in the vignette including information about group means as well as correlation direction and strength. We did not explain the concept of correlation to participants, but we did provide them with two data points to help participants conceptualize the direction and size of the effect. All groups were shown the same article, except the graph groups viewed the graph associated with their assigned group (see Figure 1, Figures S1–S4 for news article stimuli).

Prior Belief. Before presenting participants with the article, we asked them to rate their agreement with the question: "Does attending religious service increase happiness?" on a scale of 1(*Strongly disagree*) to 7 (*Strongly agree*). We collected this measure so that we could see whether participants generally had strong prior beliefs about the scenario, given prior work showing that the persuasive effects of graphs are most likely to emerge in scenarios where people have weak prior beliefs (Pandey et al., 2014).

Impression of the Article. To assess the influence of graphs on scientific reasoning about correlational relationships, participants were asked several questions after reading the article. They were first asked to answer the question "How effective is attending religious service on improving happiness?" on a scale of 1 (*Not effective at all*) to 7 (*Extremely effective*). This language was borrowed from Tal and Wansink (2016) where they ask participants "How effective is the medication?." To determine whether participants perceived causality from the correlational relationship reported in the article, we next asked "Do you think that the study shows that attending religious service causes increased happiness?," which they answered on a scale of 1 (*Not at all*) to 7 (*Absolutely*). To assess whether the presence of a graph increased the association of the described research with science, we asked, "Does the study described in the article provide a scientific basis for attending religious service to improve happiness?" which they answered on a scale of 1 (*Not at all*) to 7 (*Absolutely*). We also asked participants the question, "Do you think that therapists should list attending religious service as one of the possible ways to reduce depression symptoms in their patients?" which they responded to on a scale of 1 (*Strongly disagree*) to 7 (*Strongly agree*).

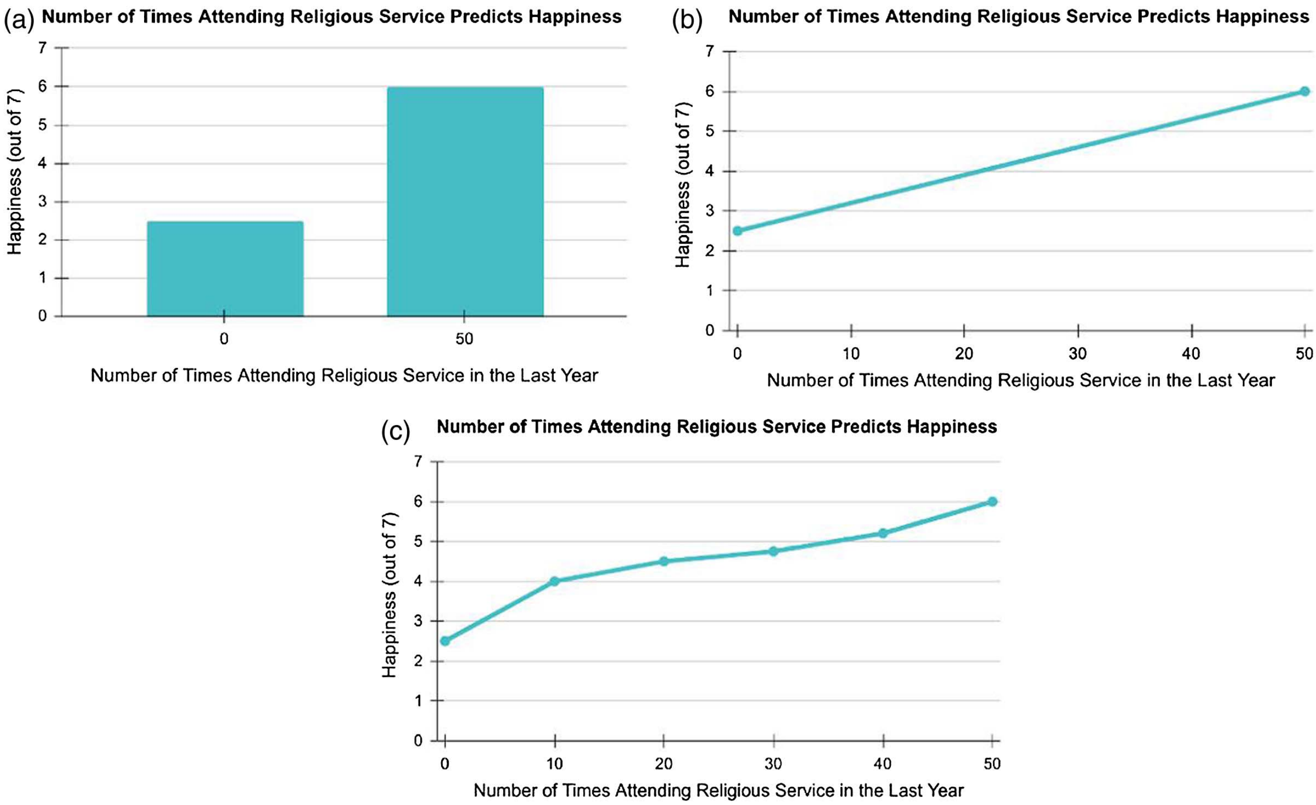
Real-Life Scenario. We also assessed whether participants would apply the conclusion of the article that going to religious service leads to increased happiness to a novel scenario. Participants were given the following vignette:

> Zahra is a new mother who has a history of moderate depression. She recently scored a 28 on the Beck Depression Inventory, which is associated with moderate depression. Zahra read this article in Scientific Citizen and is interested in the results because she does attend church sporadically and is interested in possible behavioral approaches to addressing her depression. Right now, she's free and flexible on Sunday mornings.

Participants were asked to rate on a scale of 1 (*No impact*) to 7 (*Extremely helpful*) how helpful they thought attending church more regularly would be for Zahra. They were also given information about the Beck Depression Inventory (minimal score: 0–13, mild score: 14–19, moderate score: 20–28, severe score: 29–43), and were asked to predict what Zahra's Beck score would be after regularly attending religious service for 2 months, keeping in mind that her original score was 28. Participants answered this question using a slider-scale interface from 0 to 63 with anchors at every 3 points. Lastly, they were told that Zahra's current antidepressant medication dose was 100 mg and that patients with depression typically take anywhere from 50 to 150 mg of this medication. Participants were asked to predict the dose of the medication Zahra would need after 2 months of attending religious service with a slider scale from 0 to 150 mg, with anchors at every 15 mg.

Figure 1

Stimuli Included With the Vignette for Those in the (a) Bar Graph Group, (b) 2-point Line Graph Group, and (c) 6-point Line Graph Group



Note. See the online article for the color version of this figure.

Belief in Science. We assessed belief in science with the ATS scale (Francis & Greer, 1999). The scale consisted of 10 items for which participants rated their agreement on a scale of 1 (*strongly disagree*) to 7 (*strongly agree*). These items were:

1. Science and technology cause many of the world's problems.
2. Working in a science career would be an interesting way to earn a living.
3. Science is very important for the country's development.
4. Science is very important for my community's development.
5. Money spent on science is well worth spending.
6. In my future career, I would like to use the science I learned in school.
7. The science taught in school is interesting.
8. Science is relevant to my everyday life.
9. I do not have much interest in science.
10. I would like to understand more about scientific explanations of things.

Demographics. Participants provided their age, gender, and highest level of education.

Ensuring Data Quality. To ensure that our data was of the highest quality, participants had to complete a CAPTCHA before taking the survey ("I am not a robot") and an attention check item was embedded in the ATS survey asking participants to please select *strongly agree*.

Procedure

Amazon Mechanical Turk workers were invited to participate in an online study in which they would have to read and answer questions about news articles. After deciding to participate, participants were redirected to a Qualtrics link where they had to complete a CAPTCHA and provide informed consent before proceeding to the experiment. After beginning the experiment, participants were given the prior belief item followed by the instructions:

> Scientific Citizen is a reputable news source publishing the latest research in medicine and science. In this study you will be asked to read news articles that were recently published about highly publicized findings in psychology. Please carefully read the articles before responding to the questions.

They were then randomly assigned to one of the four conditions and were presented with the corresponding news article. After reading the news article, participants were asked the article-related and applied questions. Lastly, participants completed the ATS scale, provided their basic demographic information, self-reported their effort on a scale of 1−10, and were debriefed. After finishing the survey, the participants were given a randomly generated completion code that was used to confirm their participation for compensation.

Results

Analysis Methods

Data Collection Procedure. As preregistered, we originally aimed to collect data from 400 participants. We were able to collect data from 397 participants. After fitting our models, we examined the 95% highest posterior density interval for the Condition contrasts and were dissatisfied with the precision of the estimates. Thus, we decided to collect data from an additional 400 participants in an attempt to increase the precision of our estimates (i.e., decrease the widths of the 95% highest posterior density intervals), resulting in a total sample size of 803 participants. Our sample sizes were not exactly 400 for either round of data collection due to difficulties with verifying MTurk workers' survey submissions. Ultimately, doubling our sample size resulted in minimal increases in the precision of our estimates, so we elected to halt data collection after two rounds (see Supplemental Materials to compare model precision after rounds 1 and 2 of data collection). Note that these analyses of precision focused on an earlier version of the model with non-informative priors.

Exclusion Criteria. Participants were excluded for failing an attention check item embedded into the ATS scale (n = 14) or self-reporting effort of less than or equal to 5 out of 10 (n = 7). Seven hundred and eighty-two participants remained after applying the exclusion criteria, with 198 in the text-only condition, 192 in the bar graph condition, 191 in the 2-point line graph condition, and 201 in the 6-point line graph condition.

Model Fitting. Outcome variables were analyzed using Bayesian distributional regression models with Condition, ATS, and their interaction as predictors. Condition was coded using a treatment contrast with the text-only condition as the reference group. We selected different distributional families for the models depending on the type of outcome variable being modeled. For example, ordinal variables were modeled using the cumulative logistic distributional family and continuous variables were modeled with the gaussian or exgaussian distributional families.

Models were implemented using the R-package brms (Bürkner, 2017), which converts R-style modeling syntax to Stan code (Carpenter et al., 2017). Stan is a probabilistic programming language for specifying probabilistic models and for performing approximate Bayesian inference over those models using Markov-Chain Monte-Carlo (MCMC) sampling. Four MCMC sampling chains with 2000 total iterations and 1,000 warm up iterations were run for each model. If it was indicated that further iterations were needed for model convergence, we increased the number of iterations to 5,000 with 2,500 warmup iterations. While earlier versions of the model used noninformative priors, this precluded the use of Bayes Factors, which are undefined (or approach infinity) when prior probabilities are (or approach) zero. So, to enable the use of Bayes Factors, we assigned weakly informative priors to all parameters. For slopes, we used normal (0, 1) priors and for intercepts and variance parameters, we used the default Student's t priors provided by brms (v 2.14.4). To assess the accuracy of the models, we used graphical posterior predictive checks (Gabry et al., 2019). When we noticed deviation between the predicted and observed statistics, we adjusted the distributional family accordingly (e.g., gaussian to exgaussian for highly skewed data). It should also be noted that Likert-type data were originally modeled as gaussian but were switched to ordinal regression after graphical posterior predictive checks revealed extreme divergence between observed and model-predicted outcomes.

We used Bayes factors to assess whether the inclusion of a specific predictor improved the fit of the model. Bayes Factors were computed using the "bayes_factor" function in the R-package brms (Bürkner, 2017). Three model comparisons are reported for each outcome variable, in which the alternative model is compared to the null model (i.e., the model excluding the predictor of interest). Bayes factors are reported for the overall effect of graphs compared to the text-only condition, the effect of ATS, and the overall interaction between graph presence and ATS; we do not report Bayes factors for each level of Condition (i.e., for each type of graph). Bayes factor values are interpreted qualitatively with the criteria recommended by Lee and Wagenmakers (2013). The following model comparisons were run to test the impact of Condition and ATS for each of the outcome variables (y):

> Comparison 1—H_1: Effect of Condition (C) ($y \sim C + ATS$)/H_0: No Effect of Condition ($y \sim ATS$)
>
> Comparison 2—H_1: Effect of ATS ($y \sim C + ATS$)/H_0: No Effect of ATS ($y \sim C$)
>
> Comparison 3—H_1: Interaction ($y \sim C + ATS + C \times ATS$)/$H_0$: No Interaction ($y \sim C + ATS$)

Condition-specific means and standard errors of the outcome variables are displayed in Figures 2 and 3. Note that while ATS is a continuous variable in the models, we perform a median split on ATS in our visualizations. For detailed model output, see the Supplemental Materials where we report the probabilities of direction (pd) for each effect. The probability of direction is an index of effect existence similar to a frequentist p value which conveys the proportion of the posterior (50%−100%) with the same *direction* as the median estimate (Makowski et al., 2019). Detailed statistics for each effect, including posterior median (β), 95% credible interval, and probability of direction (pd), are reported in the Supplemental Materials.

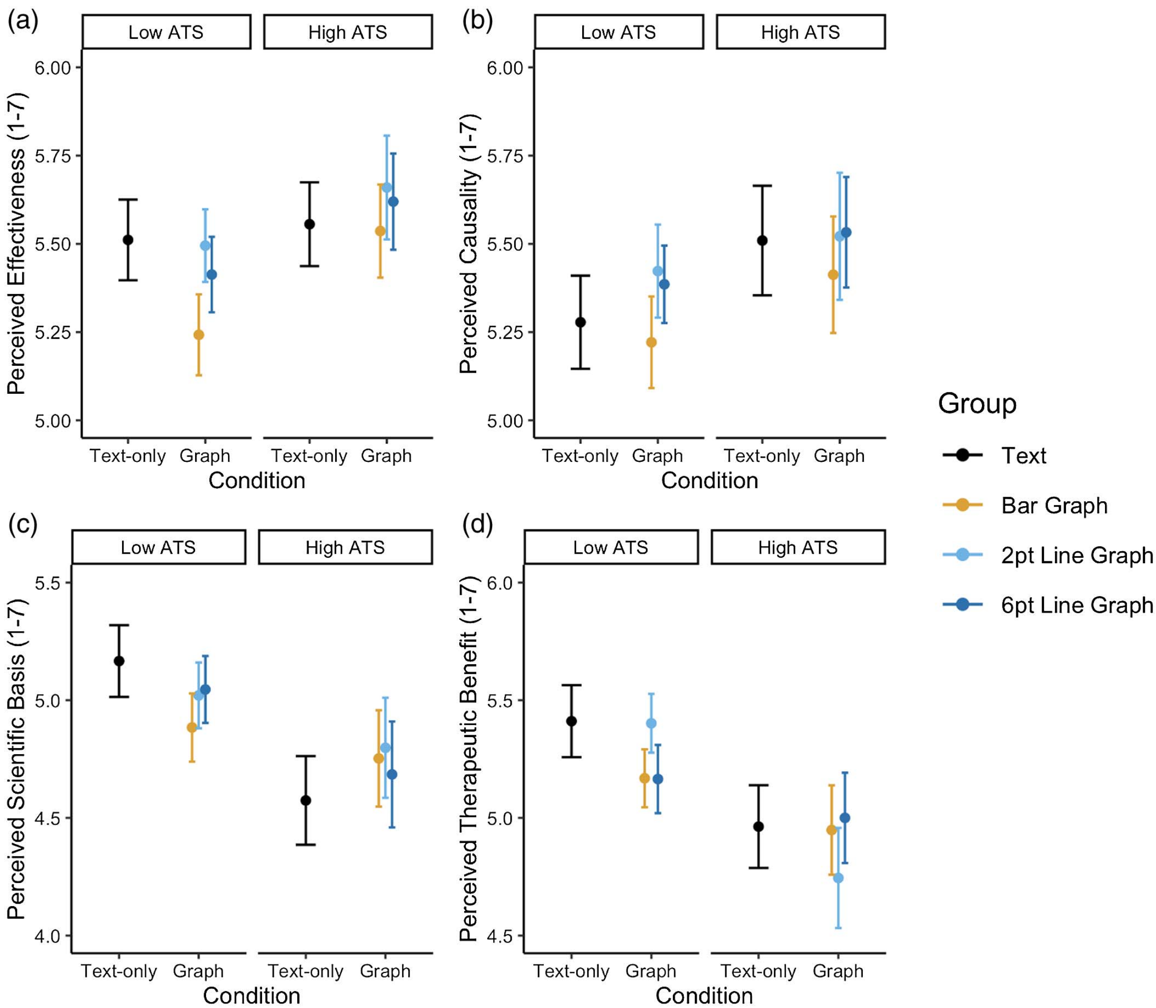
Description of the Results

Perceived Effectiveness. Bayes factor model comparisons suggest very strong evidence that there was not an overall effect of Condition on perceived effectiveness (BF = .02), no evidence as to whether ATS influenced perceived effectiveness (BF = 1), and extreme evidence that there was not an interaction between Condition and ATS (BF = .006; see Figure 2a).

Perceived Causality. Bayes factor model comparisons suggest very strong evidence that there was not an overall effect of Condition on perceived causality (BF = .01), moderate evidence for an effect of ATS (BF = 8.94), and extreme evidence that there was not an interaction between Condition and ATS (BF = .003; see Figure 2b).

Figure 2

Responses to Article-Related Questions by Condition



Note. Illustrates participant responses between text-only and graph conditions for (a) perceived effectiveness, (b) causality, and (c) scientific basis, as well as (d) whether or not participants agreed that therapists should recommend attending religious service to patients with depression. Responses are median split by attitude toward science (ATS) for the purpose of visualization, with high ATS participants in the right panel and low ATS participants in the left panel of each figure. It is important to note that ATS was modeled as a continuous variable in the analyses and that the *y*-axis for each of these plots is truncated so that the reader can clearly see each of the group means and standard errors. The true range of the outcome variable is in the label on the *y*-axis. Graphs are colored with the palette provided by (Wong, 2011). See the online article for the color version of this figure.

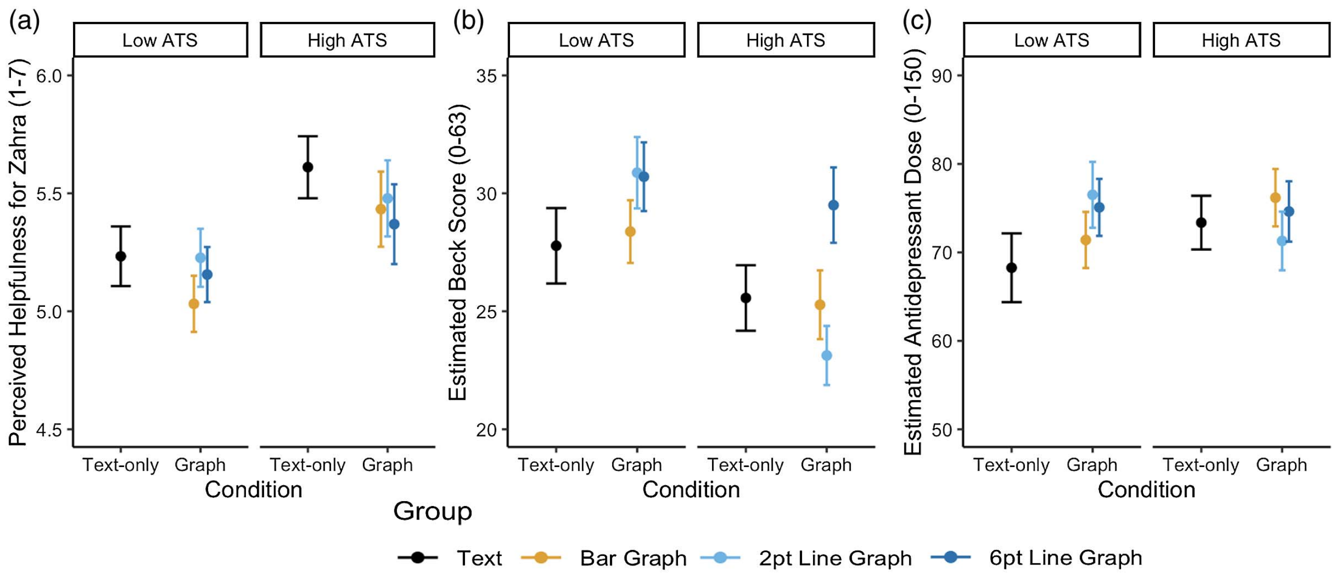
Perceived Scientific Basis. Bayes factor model comparisons suggest extreme evidence that there was not an overall effect of Condition on the perceived scientific basis for the claim (BF = .005), moderate evidence that there was no effect of ATS (BF = .12), and extreme evidence that there was not an interaction between Condition and ATS (BF = .005; see Figure 2c).

Perceived Therapeutic Benefit. Bayes factor model comparisons suggest extreme evidence that there was not an overall effect of Condition on the perceived treatment value of attending religious service (BF = .009), anecdotal evidence that there was an effect of ATS (BF = 1.44), and extreme evidence that there was not an interaction between Condition and ATS (BF = .008; see Figure 2d).

Application to a Real-Life Scenario. Bayes factor model comparisons suggest very strong evidence that there was not an overall effect of Condition on the perceived helpfulness of attending religious service (BF = .01), extreme evidence for an effect of ATS (BF > 100), and extreme evidence that there was not an interaction between Condition and ATS (BF = .002; see Figure 3a).

Participants were asked to estimate Zahra's score on the Beck inventory (Beck et al., 1961) after regularly attending religious service. These scores were rescaled by dividing by the standard deviation before fitting the model. Bayes factor model comparisons suggest extreme evidence that there was not an overall effect of Condition on estimated Beck score (BF = .002), anecdotal evidence that there was not an effect of ATS (BF = .61), and extreme evidence that there was not an interaction between Condition and ATS (BF = .002; see Figure 3b).

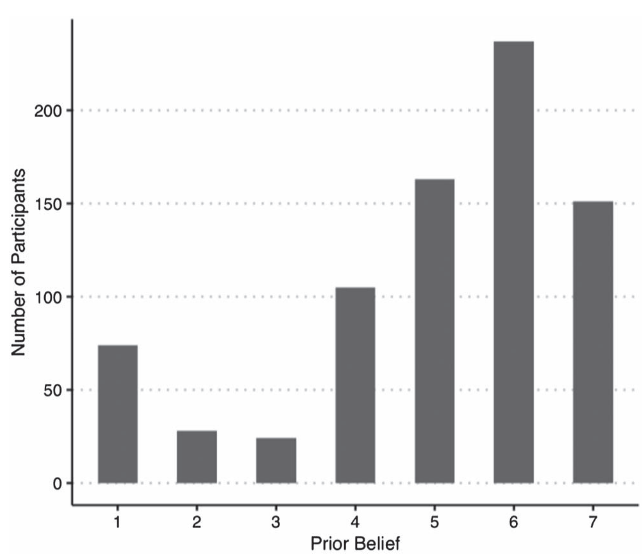
Figure 3
Responses to Applied Questions by Condition



Note. Illustrates participant responses between text-only and graph conditions for applied questions related to Zahra's depression including perceived helpfulness of attending religious service (a), estimated Beck score after regularly attending religious service (b) and estimated antidepressant dosage (c). Responses are median split by attitude toward science (ATS), with high ATS participants in the right panel and low ATS participants in the left panel of each figure. It is important to note that ATS was modeled as a continuous variable in the analyses and that the *y*-axis for each of these plots is truncated so that the reader can clearly see each of the group means and standard errors. The true range of the outcome variable is in the label on the *y*-axis. See the online article for the color version of this figure.

Lastly, participants were asked to estimate Zahra's antidepressant dosage after regularly attending religious service. These scores were rescaled by dividing by the standard deviation before fitting the model. Bayes factor model comparisons suggest very strong evidence that there was not an overall effect of Condition on estimated Beck score (BF = .03), moderate evidence that there was not an effect of ATS (BF = .22), and strong evidence that there was not an interaction between Condition and ATS (BF = .04; see Figure 3c).

Figure 4
Histogram Illustrating Prior Beliefs of Participants on the Relationship Between Attending Religious Service and Happiness



Discussion

We consistently found very strong to extreme evidence that there was no effect of graph presence, and strong to extreme evidence that there was not an interaction between the presence of a graph and ATS. Why did we fail to find an impact of graphs? It could be that the topic of religious service and happiness elicits strong prior beliefs that influenced participants' responses. As seen in Figure 4, prior belief responses were not concentrated around 4, which would have suggested a general neutral prior belief. It is well documented that prior beliefs strongly influence reasoning about data (Lord et al., 1979), and that people often engage in motivated reasoning by attempting to come to conclusions that are consistent with their biases (Kunda, 1990). Indeed, Pandey et al. (2014) found that graphs were most influential when people had neutral prior beliefs about a topic. Thus, in Experiment 2 we replicate Experiment 1 but in a more neutral context.

Experiment 2

The purpose of Experiment 2 was to see whether the impact of graphs on reasoning about correlational data would emerge in a more neutral context. Prior literature on the influence of graphs is mixed on whether or not prior beliefs influence the persuasiveness of graphs, with some researchers finding the effects for controversial topics (Heyer et al., 2020), and others suggesting that the effects only emerge when prior beliefs are neutral (Pandey et al., 2014).

In Experiment 2, participants reasoned about a made-up online news article describing a study that showed a positive correlation between the number of sporting events attended in the last year and physical fitness. We chose this topic based on Xiong et al. (2020)'s finding that participants had relatively neutral beliefs about the causal relationship between the amount of money people spent on sporting events and fitness. We hypothesized that there would be an influential effect of graph presence and interactions between graph presence and ATS for all the outcome variables.

Method

Participants

A group of 403 participants were recruited from Amazon Mechanical Turk to participate in an online study, Age, *M* (*SD*): 40.04(12.34) yrs; 43% Female, 57% Male. Workers were required to be in the U.S. to participate in our study and they were compensated $1 for their participation. We also excluded workers who had participated in Experiment 1. All procedures were determined to be exempt by the University of Michigan Institutional Review Board. The preregistration for Experiment 2 may be viewed at https://aspredicted.org/blind.php?x=si3ks7

Design

The design was the same as Experiment 1 where participants were randomly assigned to one of four conditions: text-only, bar graph, 2-point line graph, or 6-point line graph.

Materials

All materials were presented to participants via Qualtrics survey software and mirrored the materials from Experiment 1 but with a different context. The main stimuli were realistic online news articles created by the experimenter for the made-up source "Scientific Citizen." The news articles described a study showing a link between the number of sporting events attended and physical fitness:

> Researchers at University of Michigan surveyed 400 people on their basic health information and their interest in attending sporting events. Participants were asked to report the number of sporting events they had attended in the last year. The researchers also calculated each participant's physical fitness on a 7-point scale. This rating was based on their height, weight, age, gender, BMI, and body composition (amount of fat, muscle, water). Participants who did not attend any sporting events had an average physical fitness rating of 2.5, while participants attending sporting events 50 or more times had an average rating of 6. This research suggests that people who attend more sporting events tend to be more physically fit.

All groups were shown the same article, except the graph groups also viewed the graph associated with their assigned group. The trends displayed in the graphs were identical to those from Experiment 1, except the *x*-axis was renamed "number of sporting events in the last year" and the *y*-axis was renamed "physical fitness," see Figure 1, Figures S5–S9 for article stimuli.

Prior Belief. Participants were asked to rate how they agreed with the following statement: The more sporting events people attend, the more physically fit they tend to be on a scale of 1 (*Strongly Disagree*) to 7 (*Strongly Agree*).

Impression of the Article. To assess the influence of the presence of graphs on scientific reasoning about correlational relationships, participants were asked modified versions of the questions from Experiment 1. They were first asked to answer the question "How effective is attending sporting events on increasing physical fitness?" on a scale of 1 (*Not effective at all*) to 7 (*Extremely effective*). To examine whether participants perceive causality from the correlational relationship reported in the article, we next asked "Do you think that the study shows that attending sporting events causes increased physical fitness?," which they answered on a scale of 1 (*Not at all*) to 7 (*Absolutely*). To assess whether the presence of a graph increases the association of the research to "science," we asked, "Does the study described in the article provide a scientific basis for attending sporting events to increase physical fitness?" which they responded to on a scale of 1 (*Not at all*) to 7 (*Absolutely*). We also asked the question "Do you think that personal trainers should list attending sporting events as a possible method to increase physical fitness?" which participants responded to on a scale of 1 (*Strongly disagree*) to 7 (*Strongly agree*).

Real-Life Scenario. We again assessed whether participants would apply the conclusion of the article to a real-life scenario. The scenario and questions mirrored those asked to participants in Experiment 1. Participants were given the following vignette:

> Zahra is a new mother who is working on her physical fitness. She currently has a Body Mass Index (BMI) of 33, which is considered obese. Zahra read this article in Scientific Citizen and is interested in the results because she does attend sporting events sporadically and is interested in possible methods for addressing her fitness.

Participants were asked to rate on a scale of 1 (*No Impact*) to 7 (*Extremely Helpful*) how helpful they thought attending more sporting events would be for Zahra. They were also given information on Body Mass Index (BMI) scores (underweight: 15–19, normal: 20–24, overweight: 25–29, obese: 30–34, severely obese: 35–39, morbidly obese: 40–45), and were asked to predict what Zahra's BMI would be after attending a sporting event every week for 2 months, keeping in mind that her original BMI was 33. Participants answered this question using a slider-scale interface from 14 to 45 with anchors at every 3 points. Lastly, they were told that Zahra's current pant size was a size 16 and they were asked to predict her pant size after 2 months of regularly attending sporting events. They were also told that women's pant sizes typically range from a size 0–20. Participants made their responses with a slider scale from 0 to 20, with anchors at every 2 sizes.

Other Measures. ATS and demographic characteristics were the same as in Experiment 1, as well as the precautions employed to ensure data quality.

Procedure

The procedure was the same as reported in Experiment 1.

Analysis Methods

Data Collection Procedure. As preregistered, we had an original goal to collect data from 400 participants and ended up with data from 404 MTurk workers. We decided not to collect data from an additional 400 subjects given that doing so in Experiment 1 minimally increased precision of the 95% highest posterior density intervals (see Supplemental Materials).

Exclusion Criteria. Participants were excluded for failing an attention check item embedded into the ATS scale ($n = 4$) or self-reporting effort of less than or equal to 5/10 ($n = 0$). Four hundred participants remained after applying the exclusion criteria, with 100 in the text-only group, 103 in the 2-bar bar graph group, 98 in the 2-point line graph group, and 99 in the 6-point line graph group.

Model Fitting. All model-fitting and Bayes factor analysis procedures were identical to those reported in Experiment 1. For detailed model output, please see the Supplemental Materials.

Results

Perceived Effectiveness

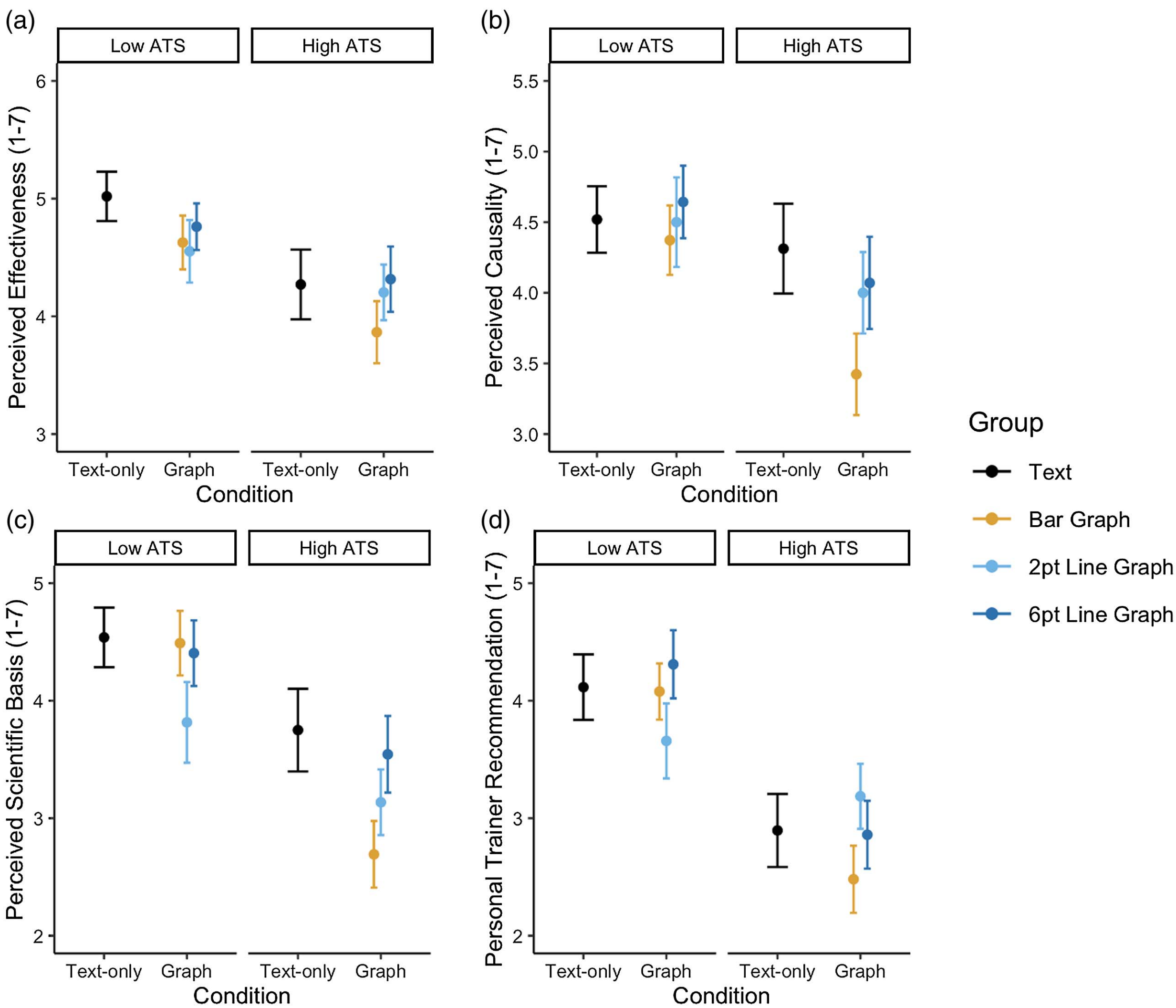
Bayes factor model comparisons suggest strong evidence that there was not an overall effect of Condition on perceived effectiveness (BF = .05), moderate evidence that there was not an effect of ATS (BF = .27), and very strong evidence that there was not an interaction between Condition and ATS (BF = .01; see Figure 5a).

Perceived Causality

Bayes factor model comparisons suggest strong evidence that there was not an overall effect of Condition on perceived causality

Figure 5

Responses to Article-Related Questions by Condition



Note. Illustrates participant responses between text-only and graph conditions for perceived effectiveness (a), causality (b), and scientific basis (c), as well as whether or not participants agreed that personal trainers should recommend attending sporting events to their overweight clients (d). Responses are median split by attitude toward science (ATS), with high ATS participants in the right panel and low ATS participants in the left panel of each figure. It is important to note that ATS was modeled as a continuous variable in the analyses and that the *y*-axis for each of these plots is truncated so that the reader can clearly see each of the group means and standard errors. The true range of the outcome variable is in the label on the *y*-axis. See the online article for the color version of this figure.

(BF = .06), anecdotal evidence for no effect of ATS (BF = .41), and very strong evidence that there was not an interaction between Condition and ATS (BF = .01; see Figure 5b).

Perceived Scientific Basis

Bayes factor model comparisons suggest moderate evidence that there was not an overall effect of Condition on perceived causality (BF = .11), moderate evidence for an effect of ATS (BF = 8.82), and very strong evidence that there was not an interaction between Condition and ATS (BF = .02; see Figure 5c).

Perceived Treatment Value

Bayes factor model comparisons suggest very strong evidence that there was not an overall effect of Condition on perceived treatment value (BF = .02), extreme evidence for an effect of ATS (BF > 100), and extreme evidence that there was not an interaction between Condition and ATS (BF = .005; see Figure 5d).

Application to a Real-Life Scenario

Bayes factor model comparisons suggest strong evidence that there was not an overall effect of Condition on perceived helpfulness of attending sporting events for Zahra (BF = .09), strong evidence for an effect of ATS (BF = 15), and very strong evidence that there was not an interaction between Condition and ATS (BF = .03; see Figure 6a).

Participants were asked to estimate Zahra's BMI after regularly attending sporting events. These scores were rescaled by dividing by the standard deviation before fitting the model. Bayes factor model comparisons suggest very strong evidence that there was not an overall effect of Condition on estimated BMI (BF = .03), moderate evidence for no effect of ATS (BF = .13), and extreme evidence that there was not an interaction between Condition and ATS (BF = .002; see Figure 6b).

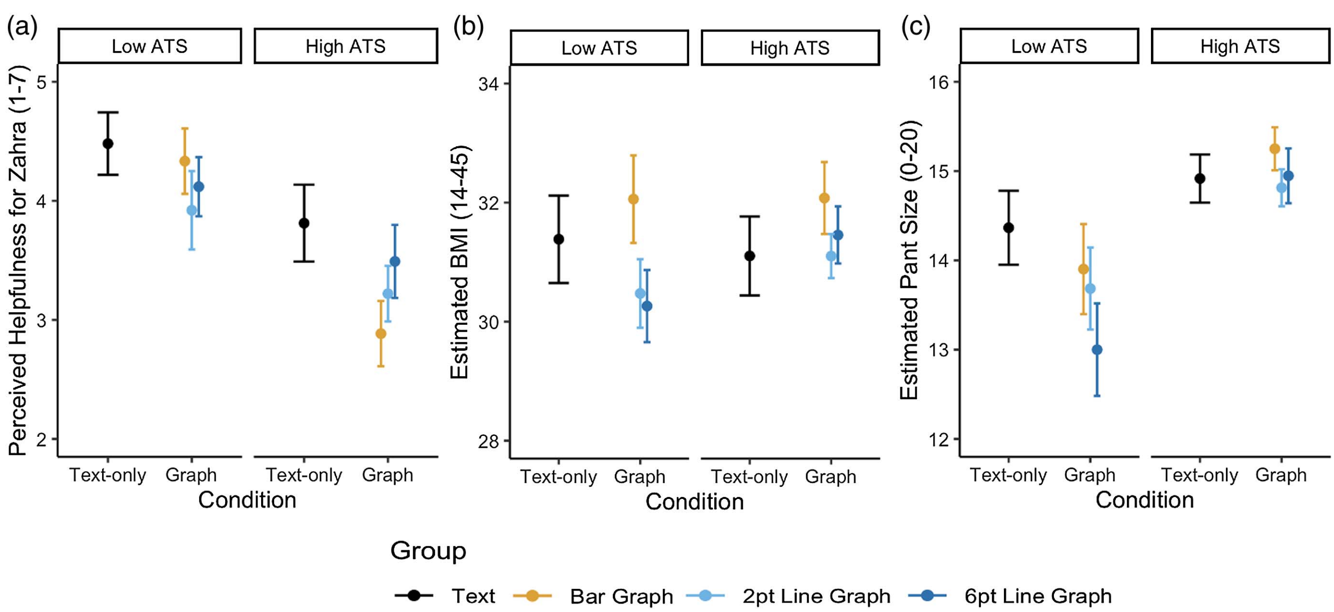
Lastly, participants were asked to estimate Zahra's pant size after regularly attending sporting events. These scores were rescaled by dividing by the standard deviation before fitting the model. Bayes factor model comparisons suggest very strong evidence that there was not an overall effect of Condition on estimated pant size (BF = .01), extreme evidence for an effect of ATS (BF > 100), and very strong evidence that there was not an interaction between Condition and ATS (BF = .01; see Figure 6d).

Discussion

We were again unable to find an influential effect of graph presence on scientific reasoning. We consistently found evidence that there was no effect of graph presence on the outcome variables and that there were no interactions between graph presence and ATS. Because we neither found an effect of graph presence nor that people with high ATS were more influenced by the presence of graphs, we decided to conduct a direct replication of Tal and Wansink (2016) since other researchers have also failed to replicate these findings in past research (Dragicevic & Jansen, 2018).

Figure 6

Responses to Applied Questions by Condition



Note. Illustrates participant responses between text-only and graph conditions for applied questions related to Zahra's weight loss including perceived helpfulness of attending sporting events (a), estimated body mass index (BMI) score after regularly attending sporting events (b) and estimated pant size (c). Responses are median split by attitude toward science (ATS), with high ATS participants in the right panel and low ATS participants in the left panel of each figure. It is important to note that ATS was modeled as a continuous variable in the analyses and that the *y*-axis for each of these plots is truncated so that the reader can clearly see each of the group means and standard errors. The true range of the outcome variable is in the label on the *y*-axis. See the online article for the color version of this figure.

Experiment 3

Results from Experiments 1 and 2 provided evidence that graphs do not influence reasoning about correlational data for all seven of our outcome variables across both experiments. This was especially surprising given that the language used for one of these outcome variables (effectiveness of X on Y) was taken directly from Tal and Wansink (2016). Considering that our hypotheses were based on the findings from Tal and Wansink (2016), and researchers have since presented evidence that casts doubt on these claims (Dragicevic & Jansen, 2018), we attempted a direct replication of Tal and Wansink (2016) to establish whether graphs actually enhance the perceived strength of a factor and whether people with high belief in science are actually more influenced by the presence of a graph.

Method

Participants

We originally intended to analyze the data using traditional null-hypothesis testing methods and calculated power a priori to determine sample size. While Tal and Wansink (2016) did not report effect sizes in their report, we used methods from Borenstein (2009) to calculate the effect size for the main effect of Group reported in their Experiment 1 (d = .54). We conducted a 95% power analysis and determined that we would need data from 182 participants to find this effect. To account for potential loss of data due to attention check failure, we collected 20% more data than needed for a total sample of 218. We ended up collecting data from 221 participants, Age, M (SD): 36.31 (9.91); 36.68% Female, 64.32% Male, on Amazon Mechanical Turk to participate in an online study. Workers had to be in the U.S. to participate in our study and they were compensated \$0.50 for their participation. All procedures were determined to be exempt by the University of Michigan Institutional Review Board. Preregistration for Experiment 3 may be viewed at https://aspredicted.org/blind.php?x=fb3fd2

Design

The experiment employed a between-subjects design where participants were randomly assigned to a text-only group (Text group) or a group that saw the text along with a bar graph of the data (Graph group).

Materials

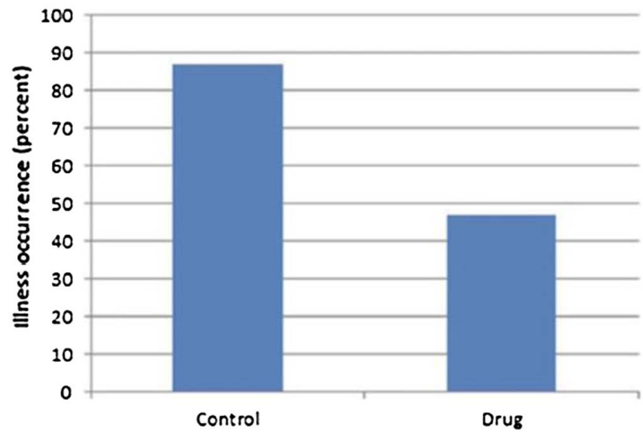
All materials were presented to participants via Qualtrics survey software. The main stimuli included either a block of text about the effectiveness of a medication (text condition):

> A large pharmaceutical company has recently developed a new drug to boost peoples' immune function. It reports that trials it conducted demonstrated a drop of 40% (from 87% to 47%) in occurrence of the common cold. It intends to market the new drug as soon as next winter, following FDA approval.

Or the block of text plus a bar graph of the data (graph condition, see Figure 7).

Effectiveness and Causal Mechanism Questions. Participants were asked "How effective is the medication?" and responded on a scale of 1 (*not at all effective*) to 9 (*very effective*). They were also asked to respond "yes" or "no" to the question "Does the medication really reduce illness?," implying a causal mechanism analogous to our prior causality questions in the context of correlational data.

Figure 7
Bar Graph From Tal and Wansink (2016) Shown to Participants in the Graph Condition



Note. See the online article for the color version of this figure.

Belief in Science. Instead of assessing belief in science with the ATS scale as in prior experiments, we use the item from Tal and Wansink (2016). Belief in science was assessed with the single item "I believe in science," where participants had to rate their agreement with the statement on a scale of 1–9.

Demographics. Participants provided their age, gender, highest level of education, and highest level of statistics education.

Ensuring Data Quality. Participants had to complete a CAPTCHA before beginning the experiment. We included 1 Likert-scale attention check item and a free response attention check. The two attention checks were mixed in with the article questions and demographic questionnaire as we did not have the ATS measure to embed an attention check item in. We also asked participants to self-report their effort on the task on a scale of 1–10, ensuring them that their response would not influence payment.

Procedure

Participants were randomly assigned to one of two conditions (text or graph). Participants viewed either text or text and a graph and were asked the article questions followed by the belief in science measure. On a separate page they completed the demographics measures. They were then debriefed and compensated.

Exclusion Criteria. Tal and Wansink (2016) did not mention data exclusion criteria in their report, so we preregistered that we would conduct our analyses both without any exclusion criteria and with our standard exclusion criteria. We found that the results did not differ as a function of whether we applied our exclusion criteria (see Supplemental Materials). Participants were excluded from the data analysis for failing either of our attention checks and/or self-reporting an effort of five or less out of ten (n = 22) yielding a sample of 199 with 100 participants in the text-only condition and 99 in the bar graph condition.

Model Fitting. All data were fitted with Bayesian regression models with Condition, Belief in Science, and their interaction as predictors. Condition was dummy coded with the text-only condition as the reference group for all analyses. The outcome variables of perceived effectiveness and whether participants really thought the medication reduced illness were modeled with Gaussian regression and Bernoulli (logistic) regression, respectively. We originally attempted to model perceived effectiveness with ordinal regression as it was a Likert-scale item; however, we found that the data were better fit by a Gaussian regression model. All other model fitting and Bayesian factor analysis procedures are identical to Experiments 1 and 2. For detailed model output, please see the Supplemental Materials.

Results

Perceived Effectiveness

Bayes factor analysis suggests anecdotal evidence for no effect of graph presence (BF = .35), extreme evidence for an effect of Belief in Science (BF > 100), and moderate evidence for no interaction between Belief in Science and graph presence (BF = .18; see Figure 8).

Does the Medication Really Reduce Illness?

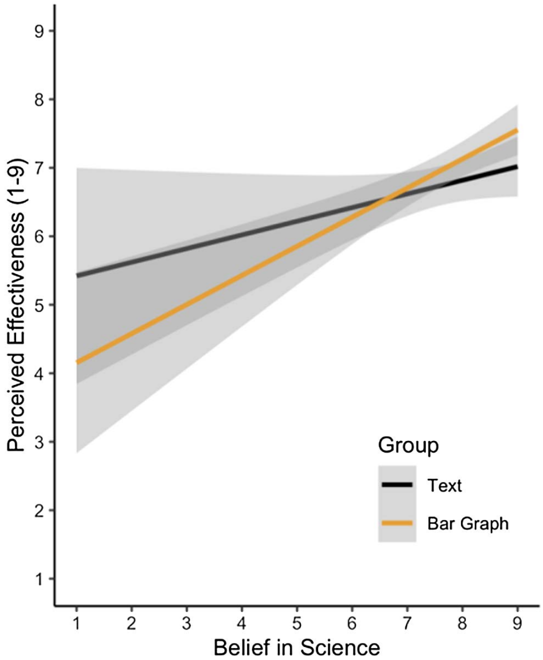
Bayes factor analysis suggests anecdotal evidence for no effect of graph presence (BF = .60), anecdotal evidence for an effect of Belief in Science (BF = 1.22), and moderate evidence for no interaction between Belief in Science and graph presence (BF = .16).

Discussion

We found anecdotal evidence that there was *no* effect of the presence of a graph on the perceived effectiveness of a medication at reducing illness. We also found moderate evidence that there was no interaction between Belief in Science and the presence of a graph. Overall, we were unable to replicate the findings from Tal and Wansink (2016) and most importantly we observed no persuasive effect of the presence of a bar graph.

Figure 8

Perceived Effectiveness as a Function of Belief in Science, Separated by Condition



Note. Gray ribbons illustrate the 95% confidence level interval for the fitted lines. See the online article for the color version of this figure.

General Discussion

Effect of Graphs on Data Interpretation

Inferring causation from correlational data can lead to faulty decision-making. To what extent does presentation format affect the likelihood of making these errors? Across two experiments, we found evidence that the presence of graphs did not impact reasoning about correlational data. We also consistently found evidence that there was not an interaction between ATS and Condition which was inconsistent with our hypotheses. Thus, in a third experiment we attempted to directly replicate the findings from Tal and Wansink (2016) in which graphs were shown to be persuasive in the context of reasoning about a drug's efficacy. We were unable to replicate their original finding that graphs had a persuasive influence on beliefs about a drug's effectiveness nor evidence of an interaction between belief in science and graph presence such that people with greater belief in science were more persuaded by the presence of a graph. Other unpublished work conducted by our lab has also failed to find an influential or persuasive effect of the presence of a graph[1]. Our failed direct replication, along with prior failed attempts at replication (Dragicevic & Jansen, 2018), suggests that at least in the context of a medication trial, graphs do not impact perception of the effectiveness of a drug and thus may not serve as a method of persuasion.

When might graphs have an impact on scientific reasoning, if ever? Graphical format influences risk perception and perceived credibility of data (Schapira et al., 2006). Things like highlighting the number of people affected (i.e., foreground) in an icon array increases risk aversion in comparison to showing icon arrays with people affected and people unaffected (i.e., foreground + background) depending on the risk probability (Okan et al., 2020). We know that altering features of graphs such as truncating the y-axis can lead to exaggerated effect sizes (Correll et al., 2020; Yang et al., 2021). And that in the judgmental forecasting literature giving participants a graph improves the forecasting of linear trends (Harvey & Bolger, 1996), but harms the forecasting of exponential trends, potentially from providing a false sense of confidence (Fansher et al., in press). Clearly, there are some circumstances where graphs do have an impact on scientific reasoning. Whether there is a general mechanism in which graphs do or do not affect scientific reasoning across domains is less clear.

Scientific Reasoning About Correlation

Although we were unable to find an influence of graphs, importantly, we found that people consistently inferred causation from correlation across Experiments 1 and 2, with an average causality rating of 5.4 on a 7-point scale in Experiment 1 and an average rating of 4.2 in Experiment 2. It is important to remember that we asked participants to rate whether or not the study *showed* that X causes Y. The presence of a correlation between two variables can certainly *provide evidence* that there may be a causal relationship between X and Y but can never prove there to be a causal relationship. It is interesting to note that the average rated causation was higher for the scenario in Experiment 1 and lower for Experiment 2. It may be that strong prior beliefs elicited by the religious service scenario led people to incorrectly infer causation more often than when reasoning about the more neutral scenario described in Experiment 2. Future work should examine how prior beliefs impact one's reasoning about correlational data. Our work further contributes to the general consensus that people are overzealous to conclude causality based on correlational data. Future research should further explore how to best display and describe correlational data given that "misunderstanding causal links can result in ineffective actions being chosen, harmful practices perpetuated, and beneficial alternatives overlooked" (Barrowman, 2014, p. 24).

Future Directions

Future work should further explore whether or not adding graphs to text increases the impact of the message. There is mounting evidence that this may not be the case (Dragicevic & Jansen, 2018; McCabe & Castel, 2008; Xiong et al., 2020). However, this topic must be studied further given its implications for science communications. Understanding the influence of graphs on reasoning and cognition is important for ethical science communications as media reports often include visualizations such as photographs, tables, and graphs. If graphs do have an undue influence, this could be taken advantage of by those with nefarious intentions or those with a particular agenda. Unfortunately, people are susceptible to believing misinformation (Lewandowsky et al., 2012; Marsh & Yang, 2018) and the nature of graphical visualizations allows viewers to be influenced by manipulating the features of graphs (Pandey et al., 2015). Thus, future work should attempt to uncover the specific conditions in which graphs are likely to affect a reader's evaluation of evidence. For example, given prior work it seems possible that the impact of graphs may only emerge in contexts where prior beliefs are neutral. Another factor that may influence the impact of graph presence is the plausibility of a causal relationship. However, if the effects of graphs only emerge under very specific conditions, researchers should also consider how generalizable these findings are to real-life applications. If the effects of graphs only emerge in very specific, artificial contexts, this work may fail to apply to everyday scientific reasoning. We encourage researchers in this area to include applied questions as it is not only important to measure how one evaluates or interprets data; it is also important to measure how reading a piece of evidence influences decision-making in real-life contexts.

Lastly, future research should attempt to uncover the cognitive mechanisms in which graphs do or do not have an impact and whether or not graphs do act as indicators of science. Tal and Wansink (2016) argue that graphs influence viewers because they act as indicators of science. They support this claim with the finding that people high in Belief in Science were most likely to increase their judgments of effectiveness of a drug in the presence of a graph. However, the current investigation found no evidence for such an interaction. One limitation of our work is that ATS and Belief in Science were relatively high across experiments, thus it is possible that ceiling effects prevented us from observing these interactions. Even if we did replicate the finding that people high in Belief in Science are most affected by the presence of graphs, this still would not provide direct evidence that graphs act as indicators of science because the concepts of believing in science and believing that graphs *are* science are much different from one another.

References

Barrowman, N. (2014). Correlation, causation, and confusion. *The New Atlantis*. https://www.thenewatlantis.com/publications/correlation-causation-and-confusion

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*(6), 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Bleske-Rechek, A., Morrison, K. M., & Heidtke, L. D. (2015). Causal inference from descriptions of experimental and non-experimental research: Public understanding of correlation-versus-causation. *The Journal of General Psychology*, *142*(1), 48–70. https://doi.org/10.1080/00221309.2014.977216

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 221–235). Sage Publications.

Bürkner, P.-C. (2017). brms: AnRPackage for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. Morgan Kaufmann.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). *Stan*: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Chua, H. F., Yates, J. F., & Shah, P. (2006). Risk avoidance: Graphs versus numbers. *Memory & Cognition*, *34*(2), 399–410. https://doi.org/10.3758/BF03193417

Correll, M., Bertini, E., & Franconeri, S. (2020). Truncating the Y-axis: Threat or menace? *ArXiv:1907.02035*. http://arxiv.org/abs/1907.02035

DeStefano, F., & Shimabukuro, T. T. (2019). The MMR vaccine and autism. *Annual Review of Virology*, *6*(1), 585–600. https://doi.org/10.1146/annurev-virology-092818-015515

Dragicevic, P., & Jansen, Y. (2018). Blinded with science or informed by charts? A replication study. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 781–790. https://doi.org/10.1109/TVCG.2017.2744298

Fagerlin, A., Zikmund-Fisher, B. J., & Ubel, P. A. (2011). Helping patients decide: Ten steps to better risk communication. *Journal of the National Cancer Institute*, *103*(19), 1436–1443. https://doi.org/10.1093/jnci/djr318

Fansher, M., Adkins, T. J., Lewis, R. L., Boduroglu, A., Lalwani, P., Quirk, M., Shah, P., & Jonides, J. (in press). How well do ordinary Americans forecast the growth of COVID-19? *Memory and Cognition*.

Francis, L. J., & Greer, J. E. (1999). Attitude toward science among secondary school pupils in Northern Ireland: Relationship with sex, age and religion. *Research in Science & Technological Education*, *17*(1), 67–74. https://doi.org/10.1080/0263514990170105

Franconeri, S., Padilla, L., Shah, P., Zacks, J., & Hullman, J. (in press). The science of visual data communication: What works. *Psychological Science in the Public Interest*.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, *182*(2), 389–402. https://doi.org/10.1111/rssa.12378

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (pp. 183–184). CRC press. https://doi.org/10.1201/b16018

Haard, J., Slater, M. D., & Long, M. (2004). Scientese and ambiguous citations in the selling of unproven medical treatments. *Health Communication*, *16*(4), 411–426. https://doi.org/10.1207/s15327027hc1604_2

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, *90*(3), 414–434. https://doi.org/10.1037/0022-0663.90.3.414

Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, *12*(1), 119–137. https://doi.org/10.1016/0169-2070(95)00634-6

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, *3*(3), 446–474. https://doi.org/10.1111/j.1756-8765.2011.01150.x

Heyer, J., Raveendranath, N. K., & Reda, K. (2020). *Pushing the (visual) narrative: The effects of prior knowledge elicitation in provocative topics* [Conference session]. Proceedings of the 2020 CHI conference on human factors in computing systems, New York, United States. https://doi.org/10.1145/3313831.3376887

Klaczynski, P. A., Fauth, J., & Gordon, D. H. (1997). Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology*, *89*(3), 470–485. https://doi.org/10.1037/0022-0663.89.3.470

Klett, L. M. (2019). Pew: People who attend church are happier than those who don't. *The Christian Post*. https://www.christianpost.com/news/pew-people-who-attend-church-are-happier-than-those-who-dont.html

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Kuruvilla, C. (2019). People who attend worship services regularly are happier than others, study suggests. *HuffPost*. https://www.huffpost.com/entry/church-attendance-happy-pew-research-center_n_5c545b34e4b0bdf0e7daac7f

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65–100. https://doi.org/10.1111/j.1551-6708.1987.tb00863.x

Lauer, C., & O'Brien, S. (2020). How people are influenced by deceptive tactics in everyday charts and graphs. *IEEE Transactions on Professional Communication*, *63*(4), 327–340. https://doi.org/10.1109/TPC.2020.3032053

Lee, M. D., & Wagenmakers, E. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *JNCI Monographs*, *1999*(25), 149–163. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024191

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, Article 2767. https://doi.org/10.3389/fpsyg.2019.02767

Marsh, E. J., & Yang, B. W. (2018). Believing things that are not true: A cognitive science perspective on misinformation. In B. G. Southwell, E. A. Thorson & L. Sheble (Eds.), *Misinformation and mass audiences* (pp. 15–34). University of Texas Press.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*(1), 343–352. https://doi.org/10.1016/j.cognition.2007.07.017

Norris, S. P., Phillips, L. M., & Korpan, C. A. (2003). University students' interpretation of media reports of science and its relationship to background knowledge, interest, and reading difficulty. *Public Understanding of Science*, *12*(2), 123–145. https://doi.org/10.1177/09636625030122001

Okan, Y., Stone, E. R., Parillo, J., Bruine de Bruin, W., & Parker, A. M. (2020). Probability size matters: The effect of foreground-only versus foreground+background graphs on risk aversion diminishes with larger probabilities. *Risk Analysis*, *40*(4), 771–788. https://doi.org/10.1111/risa.13431

Pandey, A. V., Manivannan, A., Nov, O., Satterthwaite, M., & Bertini, E. (2014). The Persuasive power of data visualization. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2211–2220. https://doi.org/10.1109/TVCG.2014.2346419

Pandey, A. V., Rall, K., Satterthwaite, M. L., Nov, O., & Bertini, E. (2015). *How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques* [Conference session]. Proceedings of the 33rd annual ACM conference on human factors in computing systems - CHI '15, New York, United States. https://doi.org/10.1145/2702123.2702608

Pew Research Center. (2019). *Religion's relationship to happiness, civic engagement and health around the world*. https://www.pewforum.org/2019/01/31/religions-relationship-to-happiness-civic-engagement-and-health-around-the-world/

Rhodes, R. E., Rodriguez, F., & Shah, P. (2014). Explaining the alluring influence of neuroscience information on scientific reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1432–1440. https://doi.org/10.1037/a0036844

Rodriguez, F., Ng, A., & Shah, P. (2016). Do college students notice errors in evidence when critically evaluating research findings? *Journal on Excellence in College Teaching*, *27*(3), 63–78.

Schapira, M. M., Nattinger, A. B., & McAuliffe, T. L. (2006). The influence of graphic format on breast cancer risk communication. *Journal of Health Communication*, *11*(6), 569–582. https://doi.org/10.1080/10810730600829916

Shah, P., Michal, A., Ibrahim, A., Rhodes, R., & Rodriguez, F. (2017). What makes everyday scientific reasoning so challenging? In B. Ross (Ed.), *Psychology of learning and motivation* (Vol. 66, pp. 251–299). Academic Press.

Smith, L. D., Best, L. A., Stubbs, D. A., Archibald, A. B., & Roberson-Nay, R. (2002). Constructing knowledge. The role of graphs and tables in hard and soft psychology. *American Psychologist*, *57*(10), 749–761. https://doi.org/10.1037/0003-066X.57.10.749

Tal, A. (2015). Looks like science, must be true: graphs and the halo of scientific truth. *Journal of Expertise*, *27*(2), Article 12.

Tal, A., & Wansink, B. (2016). Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. *Public Understanding of Science*, *25*(1), 117–125. https://doi.org/10.1177/0963662514549688

van der Linden, S. L., Leiserowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2014). How to communicate the scientific consensus on climate change: Plain facts, pie charts or metaphors? *Climatic Change*, *126*(1–2), 255–262. https://doi.org/10.1007/s10584-014-1190-4

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, *20*(3), 470–477. https://doi.org/10.1162/jocn.2008.20040

Wong, B. (2011). Color blindness. *Nature Methods*, *8*(6), 441–442. https://doi.org/10.1038/nmeth.1618

Xiong, C., Shapiro, J., Hullman, J., & Franconeri, S. (2020). Illusion of causality in visualized data. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 853–862. https://doi.org/10.1109/TVCG.2019.2934399

Yang, B. W., Restrepo, C. V., Stanley, M. L., & Marsh, E. J. (2021). Truncating bar graphs persistently misleads viewers. *Journal of Applied Research in Memory and Cognition*, *10*(2), 298–311. https://doi.org/10.1016/j.jarmac.2020.10.002

Received October 28, 2020
Revision received July 15, 2021
Accepted July 17, 2021 ■