+Model
ARTICLE IN PRESS
Journal of Applied Research in Memory and Cognition xxx (2021) xxx–xxx

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Empirical article

Truncating Bar Graphs Persistently Misleads Viewers

Brenda W. Yang*,[1]
Department of Psychology and Neuroscience, Duke University. 417 Chapel Dr., Durham, NC, 27708 US
Center for Cognitive Neuroscience, Duke University. 308 Research Dr., Durham, NC, 27710 US

Camila Vargas Restrepo[1]
Department of Psychology and Neuroscience, Duke University. 417 Chapel Dr., Durham, NC, 27708 US

Matthew L. Stanley
Department of Psychology and Neuroscience, Duke University. 417 Chapel Dr., Durham, NC, 27708 US
Center for Cognitive Neuroscience, Duke University. 308 Research Dr., Durham, NC, 27710 US

Elizabeth J. Marsh
Department of Psychology and Neuroscience, Duke University. 417 Chapel Dr., Durham, NC, 27708 US

Data visualizations and graphs are increasingly common in both scientific and mass media settings. While graphs are useful tools for communicating patterns in data, they also have the potential to mislead viewers. In five studies, we provide empirical evidence that y-axis truncation leads viewers to perceive illustrated differences as larger (i.e., a *truncation effect)*. This effect persisted after viewers were taught about the effects of y-axis truncation and was robust across participants, with 83.5% of participants across these 5 studies showing a truncation effect. We also found that individual differences in graph literacy failed to predict the size of individuals' truncation effects. PhD students in both quantitative fields and the humanities were susceptible to the truncation effect, but quantitative PhD students were slightly more resistant when no warning about truncated axes was provided. We discuss the implications of these results for the underlying mechanisms and make practical recommendations for training critical consumers and creators of graphs.

Keywords: Data visualization, Misleading graphs, Misinformation, Axis truncation, Bar graphs

General Audience Summary

News media, opinion pieces, social media, and scientific publications are full of graphs meant to communicate and persuade. Such graphs may be technically accurate in displaying correct numerical values and yet misleading because they lead people to draw inappropriate conclusions. In five studies, we investigate the practice of truncating the y-axis of bar graphs to start at a non-zero value. While this has been called one of "the worst of crimes in data visualization" by *The Economist*, it is surprisingly common in not just news and social media, but also in scientific conferences and publications. This might be because the injunction to "not truncate the axis!" may be seen as more dogmatic than data-driven. We examine how truncated graphs consistently lead people to perceive a larger difference between two quantities in five studies, and we find that 83.5% of participants across studies show a *truncation effect*. In other words, 83.5% of people in our studies judged differences illustrated by truncated bar graphs as larger than differences illustrated by graphs where the y-axis starts at 0.

Author Note.
* Correspondence concerning this article should be addressed to Brenda W. Yang, 417 Chapel Drive, Durham, NC, 27708 US.

E-mail address: brenda.yang@duke.edu (B.W. Yang).
[1] Indicates equal contribution to the present work.

Please cite this article in press as: Yang, B. W., et al. Truncating Bar Graphs Persistently Misleads Viewers. *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2020.10.002

Surprisingly, we found that the truncation effect was very persistent. People were misled by y-axis truncation even when we thoroughly explained the technique right before they rated graphs, although this warning reduced the degree to which people were misled. People with extensive experience working with data and statistics (i.e., PhD students in quantitative fields) were also susceptible to the truncation effect. Overall, our work shows the consequences of truncating bar graphs and the extent to which interventions, such as warning people, can help but are limited in their scope.

People encounter graphs in places ranging from fitness trackers to energy bills, as well as in mass media, such as broadcast television and Twitter feeds. Visualizations are often helpful, improving judgments about health care (Galesic & Garcia-Retamero, 2011; Garcia-Retamero & Galesic, 2010) and nudging people to reduce energy use (Fischer, 2008; Jensen, 2003). Graphs are invaluable tools for communicating patterns in data. For example, graphs have been deployed in COVID-19 communications to illustrate the utility of social distancing ("flattening the curve"; Roberts, 2020) and to compare properties of COVID-19 to those of other infectious diseases (Roser, Ritchie, & Ortiz-Ospina, 2020). However, there is a need to systematically examine graphs as a potential source of misinformation. Not only do graphs have the potential to be powerful, memorable, and persuasive (Newman, Garry, Bernstein, Kantner, & Lindsay, 2012; Peterson, 1983; Sargent, 2007; Standing, Conezio, & Haber, 1970), graphs are also theoretically interesting because they do not need to be factually wrong to mislead.

To wit, Figure 1 shows a bar graph appearing to communicate that taxes in the United States will increase dramatically if the Bush tax cuts expire (Shere, 2012). Closer inspection reveals that this graph exaggerates the difference between 35% and 39.5%. Additional illustrative examples of deceptive visualizations are available at osf.io/gacrj. In this paper, we examine how, under what conditions, and in what populations graphs can be misleading.

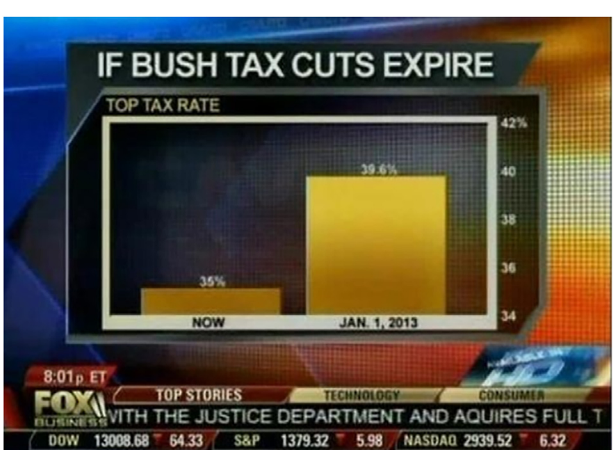

Figure 1. Example of a truncated bar graph from news media.

Studying Deceptive Visualizations

While discussions of misleading graphs are not new (e.g., Huff, 1954), empirical research on their assumed consequences is scattered across fields. Psychological research on images and visualizations suggests that graphs may increase the persuasiveness of claims, even when they provide no additional information; however, evidence on this point is mixed (Dragicevic & Jansen, 2018; Michael, Newman, Vuorre, Cumming, & Garry, 2013). Past work in the behavioral sciences also explored the cognitive processes underpinning graph perception (Carpenter & Shah, 1998; Shah & Freedman, 2011; Shah & Hoeffner, 2002), examined how people draw accurate conclusions from line graphs and scatterplots (Cleveland, Diaconis, & Mcgill, 1982; Gattis & Holyoak, 1996), and studied distortions that result from design choices (Tversky & Schiano, 1989; Zacks, Levy, Tversky, & Schiano, 1998). However, much of the work directly relevant to the effects of distorted graphs, as commonly seen in mass media outlets, comes from finance and accounting, in part because graph use is especially common in financial documents (e.g., annual reports). This largely descriptive research suggests that 20-30% of those graphs are distorted (Beattie & Jones, 1992, 1999; Beattie & Jones, 2008; Cho, Michelon, & Patten, 2012; Courtis, 1997; Penrose, 1973; Steinbart, 1989). While not experimental in methodology, this work provided early and thoughtful prescriptive guidelines for avoiding "measurement distortion" (Taylor & Anderson, 1986).

There are many categories of data visualizations, and past work described taxonomies for organizing graph types (e.g., Bertin, 1983; Tufte, 2001; Wilkinson, 2006). Here, we focus on bar graphs, a simple and powerful mapping of quantity to length found throughout scientific communications, mass media, and educational contexts. Bar graphs were the most common visualization in a randomly selected sample of articles published in *Science* in 2014 and the most represented type of visualization from news media (Borkin et al., 2013; Mogull & Stanfield, 2015). Constructing bar graphs is an explicit part of the United States' Common Core Mathematics standards starting in Grade 2 ("Draw a picture graph and a bar graph to represent a dataset with up to four categories"; Practices & Council of Chief State School Officers, 2010) and is considered a foundational category of data visualization (Few, 2012). Bar graphs are also among the easiest visualizations to create by hand and to generate in free or commercially available software.

Please cite this article in press as: Yang, B. W., et al. Truncating Bar Graphs Persistently Misleads Viewers. *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2020.10.002

Many techniques can distort the message of a bar graph. Our focus is on one of the simplest: *y-axis truncation.* This is the practice of beginning the vertical axis at a value other than the natural baseline (typically 0); it is assumed to visually exaggerate differences between graph quantities. Truncation violates a fundamental principle of data visualization, articulated by Edward Tufte and echoed by more recent recommendations: "The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented" (Cairo, 2019; Few, 2012; Tufte, 2001; Zoss, 2016). However, even though the recommendation to avoid truncation is long-standing, the practice remains common. At the time of writing, truncating the y-axis of bar graphs to illustrate small differences is the default on Microsoft Excel. Y-axis truncation is employed in mass media, textbooks, and scientific communications (osf.io/gacrj), despite being described as one of "the worst of crimes in data visualization" by *The Economist* (Leo, 2019). While some prior demonstrations suggest that y-axis truncation might impact people's perceptions, these studies have contained a single trial per condition (Pandey, Rall, Satterthwaite, Nov, & Bertini, 2015; Taylor & Anderson, 1986), small participant samples (e.g., mean n = 14.2; Witt, 2019), and/or lack of random assignment (Raschke & Steinbart, 2008). Thus, the fact that y-axis truncation reliably and systematically impacts people's perceptions of graphs remains, surprisingly, unclear.

Overview of Studies

Our first goal (Studies 1 and 2) was to provide a methodologically robust paradigm for studying how y-axis truncation affects people's judgments, with a focus on understanding the size of the *truncation effect* and allowing for further examination of possible moderators.

Given that distorted graphs are unlikely to disappear, our second goal was to explore whether we could inoculate viewers against the truncation effect, for both practical and theoretical reasons. In Studies 3–5, we investigate whether and to what extent an explanatory warning about y-axis truncation affects people's judgments. Whether or not this intervention helps has implications for whether the effect is more automatic (akin to a perceptual illusion) versus a failure to engage in controlled processing. To be clear, the warning is given prior to judging any graphs ("pre-encoding"). This pinpoints a different mechanism than warnings given after processing, as "post-encoding" warnings are often leveraged to encourage source monitoring. Our question is whether the explanatory warning increases vigilance, given that (pre-encoding) warnings that misinformation might occur slow reading (Greene, Flynn, & Loftus, 1982) and reduce susceptibility to misinformation that contradicts events (Ecker, Lewandowsky, & Tang, 2010) or prior knowledge (Marsh & Fazio, 2006). On the other hand, many visual illusions are byproducts of an adaptive visual system that uses depth and other cues to make sense of the world (Gregory, 2009) and as such cannot simply be unseen.

Our third goal was to investigate people's variability in susceptibility to the truncation effect. We include a well-regarded measure of graph literacy in all five studies (Garcia-Retamero, Cokely, Ghazal, & Joeris, 2016) to capture variation in people's experiences with visual representations of quantitative information(i.e., individual differences in graph literacy; Galesic & Garcia-Retamero, 2011). While formal training is not required to interpret graphs (Okan, Garcia-Retamero, Galesic, & Cokely, 2012), it is thought to be useful when mapping spatial information to meaning requires knowing arbitrary conventions. However, to preview, we did not find evidence for a systematic relationship between graph literacy and the truncation effect. In response, Study 5 took a different approach, exploring expertise. We reasoned that individuals in PhD programs in quantitative fields would have considerable experience creating and interpreting graphs, and thus, this population might not be susceptible to the truncation effect. We explore this in Study 5.

Study 1

Does y-axis truncation systematically affect people's judgments of the differences between quantities? In Study 1, we develop a novel and adaptable paradigm for studying misleading graphs to explore this foundational question. We also predicted that people with higher levels of graph literacy would be less susceptible to exaggerations from y-axis truncation. This would be consistent with cognitive and education research exploring processes underpinning graph interpretation (Carpenter & Shah, 1998; Shah & Hoeffner, 2002) and arguing for the need for explicit instruction to establish competence in graph comprehension (Glazer, 2011).

Method

Participants. We conducted Study 1 using Amazon Mechanical Turk (MTurk) and recruited 24 participants (9 women, 1 non-binary gender; M_{age} = 32 years, SD_{age} = 9.48). In this study, 62% of participants reported having at least a Bachelor's degree. Participants for this and all subsequent studies were recruited from the United States with previous task approval rates of least 85%. We chose a relatively small sample size, because Study 1 was exploratory, and this was the first time these stimuli and procedures were used. Analyses were not run until data collection was complete. Procedures in this and all following studies were approved by the Duke University IRB.

Materials. We made bar graphs communicating information about a range of topics, such as public health, geography, and technology. We note this divergence from materials used in previously published work, which was typically limited to financial information or presented graphs that used the same content framing across multiple trials. We aimed to balance methodological rigor with ecological validity in developing materials. The graphs ranged from three to five bars. Critically, two versions of each graph were created: one where the y-axis was truncated and one where the y-axis started at zero. Figure 2 shows a sample pair. The materials used for this and subsequent studies are available at osf.io/ytq3h.

We generated truncated bar graphs with similar levels of visual deception, a novel contribution of the present studies. Specifically, we quantified deceptiveness using an established

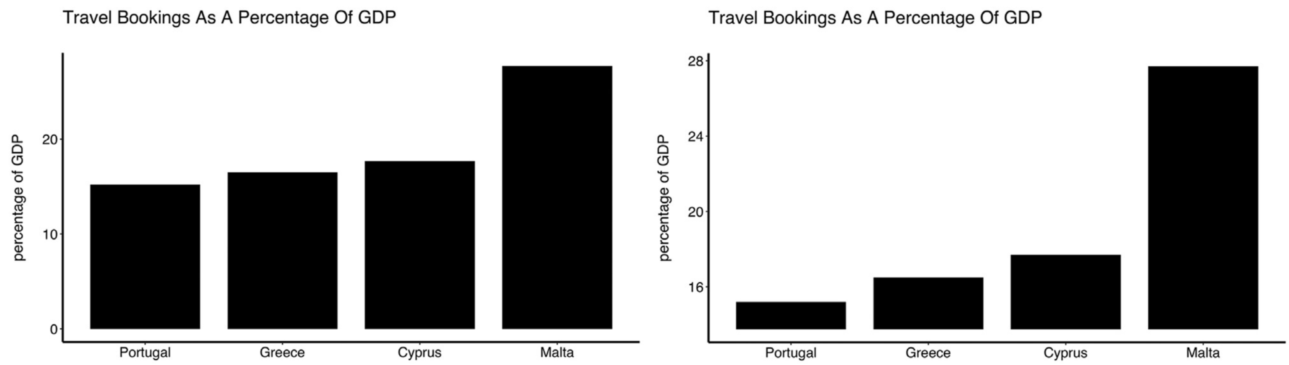

Figure. 2. Sample trial of a control graph (left), and sample trial of a truncated graph (right).

quantitative measure: the graph discrepancy index or GDI (Mather, Mather, & Ramsay, 2005; Steinbart, 1989; Taylor & Anderson, 1986; Tufte, 2001). The GDI measures the extent to which a visualization distorts the underlying numerical ratio. We created control graphs with y-axes beginning at 0 and truncated graphs with a GDI of 500. This process is described in the Supplemental Information.

We included a 10-item measure of graph literacy in this and subsequent studies, previously developed and validated by Garcia-Retamero et al., 2016. This measure asks participants to self-report their ability to interpret graphs. It is similarly reliable, robust, and valid to a much longer, objective measure which directly tests participants' ability to interpret graphs correctly. In the 10-item version implemented in our studies, participants provided ratings on 6-point scales. For example, three representative items were, "How good are you at working with bar charts?", "Are graphs easier to understand than numbers?", and "How often do you find graphical information to be useful?" The final score is a sum of all items, with higher scores indicating greater graph literacy.

Design and procedure. After providing informed consent, we told participants they would see a series of graphs and would be asked about information presented in them. Participants were instructed to look at the graphs as they would if they encountered them in a newspaper or magazine article.

After completing a sample trial, participants saw one graph at a time, proceeding at their own pace. On the same screen as the graph, participants were instructed to make a subjective comparison between two of the values represented in the bars on a scale from 1 (*not at all different*) to 7 (*extremely different*). A midpoint anchor (*moderately different*) was provided. Our choice of a relative judgment task was deliberate. We designed this task to align with the judgment tasks implied by mass media contexts such as TV or newspapers, which is often to make a relative comparison (e.g., "Taxes are *much* higher than they used to be") rather than an absolute one (e.g.,"Taxes are 12% higher than they used to be").

In Study 1, we implemented a within-subject manipulation: each participant saw 20 control bar graphs and 20 truncated bar graphs. Participants were not given information on the nature of the manipulation, and the order of the graphs was randomized. Two graph sets were created for counterbalancing, such that the two sets represented the same information but differed in which graphs were and were not truncated.

After rating 40 graphs, participants completed the 10-item graph literacy assessment described above (Garcia-Retamero et al., 2016) and answered demographic questions (age, first language, and education). Participants reported their education by selecting the highest level reached out of the following options: some high school, graduated high school or G.E.D., some college, in college, Associate's degree, Bachelor's degree, some graduate school courses, graduate degree. Finally, participants were debriefed and asked questions about their study experience; see osf.io/ytq3h for complete participant instructions for all studies.

Results

The truncation effect. First, we examined how the control versus truncated bar graph manipulation affected participant ratings of how different two quantities are (i.e., *graph ratings*). As hypothesized, we found that people rated the differences depicted by truncated graphs as larger than those depicted by control graphs: $M_{\text{control}} = 3.70$, $SD_{\text{control}} = 0.41$; $M_{\text{truncated}} = 4.78$, $SD_{\text{truncated}} = 0.63$. A paired t-test revealed that the judged difference between control graphs *versus* truncated graphs was statistically significant: $t(23) = 14.25$, $p < .0001$, 95% CI of the difference [.92, 1.24], Cohen's $d = 1.26$ [0.63, 1.90]. Figure 3a illustrates these results using modified raincloud plots, depicting means with correlation- and difference-adjusted 95% CIs (Cousineau, 2017), individual participants (2 points per participant), and distributions for both conditions side by side (Allen, Poggiali, Whitaker, Marshall, & Kievit, 2018).

Thus, when presented with truncated y-axes rather than y-axes beginning at zero, participants judged differences between bars to be substantially larger. This effect was consistent across participants: all 24 participants rated differences depicted by truncated graphs as larger than differences depicted by control graphs. We call this effect of axis truncation on subjective judgments of differences the *truncation effect*.

Graph literacy. Next, we investigated the role of graph literacy in predicting the truncation effect. Graph literacy scores ranged from 35 to 55 in Study 1 ($M = 44.08$, $SD = 5.31$). The possible range of the scores was 10–60. Table S1 in the Supplementary Information (SI) shows descriptive statistics for graph

Please cite this article in press as: Yang, B. W., et al. Truncating Bar Graphs Persistently Misleads Viewers. *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2020.10.002

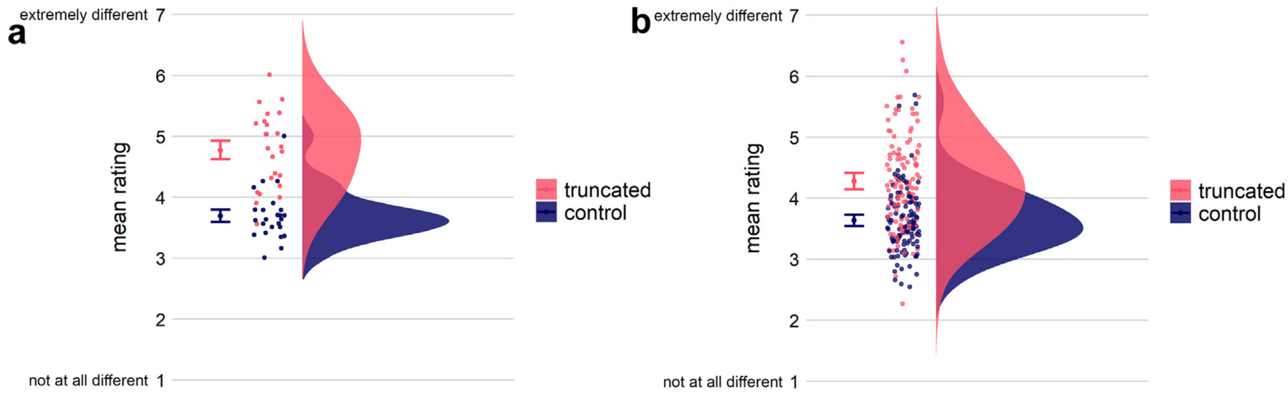

Figure 3. Raincloud plots for Study 1 (a) and Study 2 (b). These raincloud plots (Allen et al., 2018) depict average participant ratings for truncated and control graphs, respectively. Error bars reflect correlation- and difference-adjusted 95% confidence intervals of the means (Cousineau, 2017). The truncated versus control graphs variable was manipulated within-subjects; each participant is represented by two points. In this Study 2, all participants received an explanatory warning.

literacy in all studies. We found that graph literacy did not significantly predict the judged difference between bars for control versus truncated graphs: $F(1, 22) = 0.03$, $p = .87$. Figure S5 depicts this null relationship for Study 1 and the studies that follow. Thus, contrary to our hypothesis, graph literacy did not predict the size of the truncation effect. We did not find consistent relationships between graph literacy and the truncation effect in the studies that follow. Thus, in the interest of brevity, we report statistical details relating to graph literacy for subsequent studies in SI. We discuss this surprising finding in the General Discussion and have created an interactive visualization plotting the size of the truncation effect and participants' graph literacy scores at tinyurl.com/yblunbyv.

Study 2

Study 1 established a paradigm for examining the effects of y-axis truncation within bar graphs, providing a useful paradigm for studying deceptive graphs. In Study 2, we investigate whether providing an explicit explanation of y-axis truncation would reduce or eliminate the truncation effect. Clarifying what could reduce the size of the truncation effect has consequences for how misleading information might be flagged online on social networks (e.g., Clayton et al., 2019). We expected that explicit warnings about y-axis truncation would give participants the information needed to identify truncated graphs and adjust their judgments accordingly.

Method

Participants. We recruited 109 MTurk workers (50 women, 1 non-binary gender; $M_{\text{age}} = 35.07$, $SD_{\text{age}} = 11.19$; no exclusions). Because we hypothesized that a warning would reduce the truncation effect, we determined our sample size *a priori* based on a power analysis predicting a small-to-medium truncation effect size of 0.35, an alpha of 0.05, and power of 0.90, which estimated a required sample size of at least 88 participants. We did not analyze results until data collection was complete. Fifty-nine percent of participants reported having at least a Bachelor's degree. The sample also included a range of self-reported graph literacy (see Table S1 in Supplementary Information).

Materials. We used the same set of 40 bar graphs used in Study 1. We assessed graph literacy using the same 10-item scale that was used in Study 1.

Procedure. The procedure for Study 2 was very similar to Study 1. All participants saw 20 control bar graphs (where the y-axis started at zero) as well as 20 truncated bar graphs (where the y-axis did not start at zero). However, in Study 2, all participants read an explanatory warning and identified an example of a misleading graph.

This warning described y-axis truncation, provided an example, and stated that some graphs were created to be misleading. We note that, in this participant-facing context, we chose to use the terminology "misleading graph" rather than "graph with a truncated y-axis" with the reasoning that this term may be more accessible to a wide audience, and that a non-neutral label may provide more motivation to attend to the truncated graphs. As a part of the explanatory warning, participants also saw a truncated and a non-truncated version of the same graph on the same page and were asked to indicate which of the two graphs had been designed to be misleading. Regardless of performance, participants were given feedback. Eighty-eight percent of participants answered this question correctly prior to feedback.

After the warning and a sample trial, participants completed 40 trials where they made judgments about the relative differences in quantities shown by truncated and control graphs (1 = *not at all different*, 7 = *extremely different*). As in Study 1, the graphs were shown in random order and were counterbalanced. Excluding the 13 (12%) participants who did not initially answer the training exercise correctly does not change the pattern of results or conclusions drawn.

After rating the 40 graphs, participants completed the self-report graph literacy assessment used in Study 1, answered demographic questions, and answered debriefing questions.

Results

Truncation effect. We found again that people rated differences between quantities depicted by truncated bar graphs ($M = 4.28$, $SD = 0.78$) as larger than those depicted by control bar graphs ($M = 3.64$, $SD = 0.55$), as shown in Figure 3b. A paired t-test revealed this difference as statistically significant: $t(108) =$

10.64, $p < .0001$, 95% CI of the difference [0.52, 0.76], Cohen's $d = 0.54$ [0.27, 0.82]. These results suggest that despite the initial explanation and warning given, participants rated the differences in the truncated bar graph condition as larger, on average, than the differences shown by control bar graphs.

Eighty-five percent of participants showed a truncation effect in the expected direction (i.e., for 93 of our 109 participants, their ratings of comparisons shown by truncated graphs were larger than their ratings of comparisons shown by graphs without truncated vertical axes). Thus, the explicit warning immediately preceding judgments of graphs was ineffective at erasing the truncation effect.

Study 3

The results of Study 2 were surprising in that an explicit warning did not eliminate the truncation effect. To further investigate this, in Study 3, we directly manipulate in a single experiment whether participants are given an explanatory warning about y-axis truncation. Doing so allows us to directly compare the effects of having an explanatory warning or not on the truncation effect.

Method

Participants. We recruited 119 MTurk workers (49 women, 1 non-binary gender; $M_{\text{age}} = 32.27$ years, $SD_{\text{age}} = 7.81$; no exclusions). We determined our sample size by aiming to recruit a similar sample size to Study 2 as well as resource constraints for participant compensation. Seventy percent of participants reported having at least a Bachelor's degree. Results were not analyzed until all data were collected.

Materials. We used the set of 40 bar graphs used in the previous experiments. The 10-item measure of graph literacy was identical to those used in previous studies.

Procedure. Study 3 had a 2 (graph type: control, truncated) × 2 (warning, no warning) mixed design. Graph type was manipulated within-subject, while warning was manipulated between-subjects. Thus, all participants saw 20 control bar graphs (where the y-axis started at 0) and 20 truncated bar graphs (where the y-axis did not start at 0). Participants in the explanatory warning condition were given the same set of instructions, exercise, and feedback described in Study 2. Of the participants in the warning condition, 84.9% answered the training exercise question correctly. Excluding the 10 participants who did not answer the training exercise correctly initially does not change the main conclusions drawn.

As in previous studies, participants completed the graph literacy assessment and demographic questions at the end of the study.

Results

To preview, Figure 4 summarizes our primary results of interest: we replicated a truncation effect and find that it is reduced but still present when an explanatory warning is given.

The truncation effect. We first replicated our central effect of interest: the truncation effect. As in Studies 1 and 2, we found that truncated bar graphs led to exaggerated ratings of differences, compared to control bar graphs: $M_{\text{control}} = 3.86$, $SD = 0.81$; $M_{\text{truncated}} = 4.56$, $SD = 0.90$. This main effect was statistically significant: $t(118) = 10.70$, $p < .0001$, 95% CI of the difference [0.57, 0.83], Cohen's $d = 0.52$ [0.26, 0.78]. This effect was consistent: 84.0% (100 out of 119) participants showed a truncation effect in the expected direction.

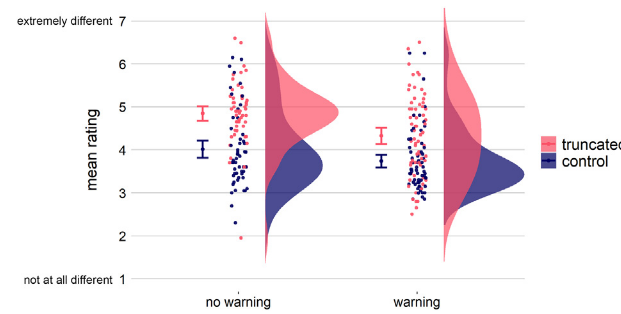

Figure 4. Raincloud plot for Study 3. Error bars reflect correlation- and difference-adjusted 95% confidence intervals of the means and points represent each participant twice. The truncated versus control graphs variable was manipulated within-subject.

Next, we examined the effects of the explanatory warning and graph type on graph ratings. We hypothesized that an explanatory warning would reduce the size of the truncation effect by lowering ratings of truncated graphs. We computed a linear mixed effects model with graph type (0 = *control*, 1 = *truncated*) and warning condition (0 = *no warning*, 1 = *warning*) as binary fixed factors. Participants' ratings of the differences depicted by bar graphs was the outcome variable, and participant and item were included as random effects. We found statistically significant effects of graph type and warning condition, as well as a significant interaction between graph type and warning (Table 1). We note that the intercept in this model (4.04) corresponds to ratings for control graphs with no explanatory warning given, acting as a theoretical and practical baseline for ratings on the 7-point scale.

Effect of an explanatory warning. To further examine the interaction between graph type and warning, we computed pairwise contrasts of graph type and warning condition from estimated marginal means derived from the linear mixed effects model described in Table 1. We implemented these analyses using the R package *emmeans* with Satterthwaite approximations to degrees of freedom (to be consistent with the approach implemented in the linear model) for Study 3 and the studies that follow (Lenth, 2019).

We first computed contrasts of graph type for each level of warning condition (no warning and warning), followed by contrasts of warning condition for each level of graph type (control and truncated). We found that in both no warning ($estimate_{\text{diff}} = 0.79$, 95% CI [0.69, 0.88], $SE = 0.05$, $p < .0001$) and warning ($estimate_{\text{diff}} = 0.63$, $SE = 0.05$, $p < .0001$ conditions, truncated graphs were rated higher than control graphs. That is, there was a robust truncation effect in both warning and no warning conditions. We also found that an explanatory warning lowered ratings for both control($estimate_{\text{diff}} = 0.32$, 95% CI [0.04, 0.60], $SE = 0.14$, $p = .03$) and truncated ($estimate_{\text{diff}} = 0.48$, 95% CI

Please cite this article in press as: Yang, B. W., et al. Truncating Bar Graphs Persistently Misleads Viewers. *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2020.10.002

Table 1
Results for Study 3 from a Linear Mixed Effects Model Predicting Graph Ratings.

	Estimate (b)	95% CI of the Estimate	SE	$\|t\|$	p
Intercept	4.04	[3.67, 4.44]	0.20	20.07	< .0001
Graph Type (Control or Truncated)	0.79	[0.68, 0.89]	0.05	15.50	< .0001
Warning Condition (No Warning or Warning)	−0.32	[−0.61, −0.04]	0.14	2.21	.03
Graph Type × Warning Condition Interaction	−0.16	[−0.29, −0.03]	0.07	2.31	.02

Note. Both experimental conditions (graph type and warning or no warning) were dummy-coded. No Warning served as the reference group for the between-subjects manipulation. Control graphs served as the reference group for the within-subject manipulation.

[0.19, 0.76], SE = .14, p = .0009) graphs. Finally, we compared these contrasts statistically to quantify the interaction found in Table 1. We found that a warning reduced graph ratings for truncated graphs more than for control graphs by an estimate of 0.16 (7-point scale), 95% CI: [0.02, 0.29], SE = 0.07, p = .02.

Thus, our results were consistent with the hypothesis that an explanatory warning reduces the truncated effect by reducing ratings for truncated graphs. However, these results suggest that an explanatory warning also led to an overall decrease in graph ratings for both types of graphs, which may indicate increased caution when rating all graphs. The truncation effect was persistent even among those who received an explicit warning.

Study 4

In Study 3, we found that providing an explanatory warning before participants rated control and truncated bar graphs reduced but did not eliminate the truncation effect. In the world, however, an explicit warning will rarely immediately precede graphs with truncated vertical axes. Here, we extend the findings of Study 3 by asking participants to provide judgments about a new set of bar graphs after a 1-day delay. The purpose in doing so was to examine whether the effects of the explicit warning on the first day will extend to the next day.

Method

Participants. A total of 157 participants (53 women, M_{age} = 33.92, SD_{age} = 9.97; no exclusions) completed both sessions of Study 4. Seventy-four percent reported having at least a Bachelor's degree. Most (90%) participants who completed Session 1 also completed Session 2. In choosing our sample size, we considered results of an *a priori* power analysis based on the size of the truncation effect given a warning (Study 3), potential attrition between sessions, and resource constraints. Data were not analyzed until all 157 participants completed the study.

Materials. We used the set of 40 bar graphs used in the previous experiments and created an additional set of 40 bar graphs as stimuli for the additional timepoint. The new set was created following the same guidelines and procedures used to create the original set. These two sets were counter-balanced across the two experimental sessions to account for potential effects of materials. The 10-item measure of graph literacy was identical to those used in previous studies.

Procedure. Study 4 had a 2 (graph type: control, truncated) × 2 (warning condition: warning, no warning in the first session) × 2 (session: first, second) mixed design. Graph type and session were within-subject factors, while we manipulated explanatory warning between subjects.

The first session of Study 4 was identical to Study 3, with half of participants receiving a warning. Most participants (76.9%) in the warning condition answered the training exercise question correctly. As before, all participants in the warning condition received feedback, either affirming their correct answer or providing corrective feedback. Excluding the participants who did not answer the training exercise correctly initially does not change the pattern of results or conclusions drawn.

One day after the first session, we asked participants to make judgments about 40 additional bar graphs that were new to the participants. Instructions at the beginning of the second session indicated that participants would look at graphs similar to the ones they had seen the day before; no explanatory warnings were given. Participants completed the graph literacy measure and demographic questions at the end of the second session.

Results

The truncation effect. We replicated the truncation effect obtained in the previous studies: average ratings of differences were higher for truncated graphs than for control graphs: M_{control} = 3.78, SD = 0.99; $M_{\text{truncated}}$ = 4.59, SD = 0.90. This main effect was statistically significant: t(156) = 15.97, p < .0001, 95% CI of the difference [0.72, 0.92], Cohen's d = 0.53 [0.31, 0.76]. Most participants showed an overall truncation effect: 88.5% of participants in Session 1 and 85.4% in Session 2.

Next, we computed a linear mixed-effects model with graph type (0 = *control*, 1 = *truncated*), warning condition (0 = *no warning*, 1 = *warning*), and timepoint (0 = *session 1*, 1 = *session 2*) as binary fixed factors. The outcome variable was graph ratings; participant and item were modeled as random effects. As expected from previous studies, we found a statistically significant effect of graph type and a statistically significant interaction between graph type and warning condition, replicating the truncation effect and the finding that that an explanatory warning leads to participants rating truncated graphs lower, compared to those not given an explanatory warning (Table 2).

We hypothesized that the protective effect of an explanatory warning would be reduced after 24 h, such that ratings for truncated graphs would be higher in the second session compared

Please cite this article in press as: Yang, B. W., et al. Truncating Bar Graphs Persistently Misleads Viewers. *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2020.10.002

Table 2
Results for Study 4 from a Linear Mixed Effects Model Predicting Graph Ratings.

	Estimate (*b*)	95% CI of the Estimate	*SE*	|*t*|	*p*
Intercept	3.76	[3.46, 4.08]	0.16	22.79	< .0001
Graph Type (Control or Truncated)	1.15	[1.06, 1.24]	0.05	24.96	< .0001
Warning Condition (No Warning or Warning)	0.19	[−0.07, 0.49]	0.14	1.37	.17
Timepoint (Session 1 or Session 2)	−0.08	[−0.17, 0.01]	0.05	1.79	.07
Graph Type × Warning Condition	−0.59	[−0.72, −0.46]	0.07	9.03	< .0001
Timepoint × Graph Type	−0.15	[−0.27, −0.02]	0.07	2.24	.03
Timepoint × Warning Condition	−0.15	[−0.28, −0.02]	0.07	2.35	.02
Timepoint × Graph Type × Warning Condition	0.13	[−0.06, 0.30]	0.09	1.35	.18

Note. Experimental conditions (graph type, warning or no warning, session 1 or session 2) were dummy-coded. No Warning and Session 1 served as the reference groups for the between-subject manipulations. Control graphs served as the reference group for the within-subject manipulation.

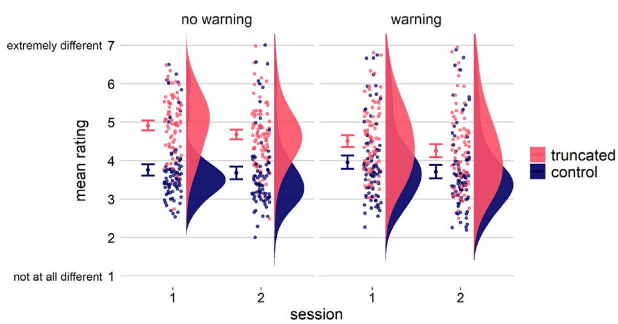

Figure 5. Raincloud Plot for Study 4. Error bars reflect correlation- and difference-adjusted 95% confidence intervals of the means. Points represent each participant twice. The session (1 and 2) and graph type (truncated and control) variables were within-subject manipulations. Cohen's *d* and 95% CIs from left to right: No Warning_1 = 0.69 [0.37, 1.02], No Warning_2 = 0.62 [0.30, 0.94]; Warning_1 = 0.42 [0.10, 0.74], Warning_2 = 0.40 [0.08, 0.72]

to the first in the warning condition. We did not find positive evidence for this hypothesis. The 3-way interaction between timepoint, warning condition, and graph type was not statistically significant ($t = 1.35$, $p = .18$). Indeed, visual inspection of Figure 5 suggests results contrary to our hypothesis: that ratings for graphs decreased at the second session. Next, we investigate each of the statistically significant 2-way interactions.

Graph type and warning condition. First, we investigated the interaction between graph type and warning condition, collapsing across timepoints. We found that in both no warning ($estimate_{\text{diff}}$ = 1.08 [1.02, 1.14], *SE* = 0.03, $p < .0001$) and warning ($estimate_{\text{diff}}$ = 0.55, 95% CI [0.49, 0.62], *SE* = 0.03, $p < .0001$) conditions, ratings for truncated graphs were higher than control graphs, consistent with the results of Study 3. Next, we computed the effect of an explanatory warning for both control and truncated graphs. As before, we found evidence that an explanatory warning lowered participants' ratings for truncated graphs: $estimate_{\text{diff}}$ = 0.41, 95% CI [0.14, 0.68], *SE* = 0.14, $p = .002$. However, the contrast between no warning and warning conditions for control graphs was not statistically significant: 95% CI [-0.38, 0.15], $p = .40$. Consistent with the statistically significant interaction between graph type and warning condition, we found that an explanatory warning lowered ratings for truncated graphs more than control graphs by an estimate of 0.53 (7-point scale; 95% CI [0.44, 0.62], *SE* = 0.05, $p < .0001$). These results suggest that an explanatory warning reduced the size of the truncation effect by lowering ratings for truncated graphs specifically.

Timepoint and graph type. To explore the interaction between timepoint and graph type, we computed contrasts of timepoint for each level of graph type (control and truncated), averaging over warning condition. We found that for both control (*estimate of the difference* = 0.16, 95% CI [0.10, 0.22], *SE* = 0.03, $p < .0001$) and truncated ($estimate_{\text{diff}}$ = 0.24, 95% CI [0.18, 0.31], *SE* = 0.03, $p < .0001$) graphs, participants' graph ratings were lower in the second session, compared to the first session. However, when comparing these contrasts, we found that lower ratings for session 2 compared to session 1 did not differ statistically as a function of graph type at an alpha of .05 (*SE* = 0.05, $p = .07$). We note that the direction of the difference was such that the decrease in graph ratings in the second session was larger for truncated graphs compared to control graphs, 95% CI [-0.007, 0.17]. Thus, we did not find further positive evidence for an interaction between graph type and timepoint.

Timepoint and warning condition. Finally, to explore the interaction between timepoint and warning condition, we computed contrasts of timepoint for the no warning and warning conditions, averaging over graph condition. We found that in both the no warning ($estimate_{\text{diff}}$ = 0.16. 95% CI [0.09, 0.22], *SE* = 0.03, $p < .0001$) and warning conditions ($estimate_{\text{diff}}$ = 0.25, 95% CI [0.18, 0.31], *SE* = 0.03, $p < .0001$), participants' graph ratings were reduced in the second session, compared to the first session. Comparing these contrasts statistically, we found that this decrease in graph ratings for the second timepoint was larger for participants who received a warning, compared to those who did not receive a warning: $estimate_{\text{diff}}$ = 0.09, 95% CI [0.00, 0.18], *SE* = 0.05, $p = .05$.

Overall, we found consistent evidence of a truncation effect with and without warning and at both timepoints. We also replicated the protective effect of an explanatory warning. We failed to find positive evidence that the protective effect of an explanatory warning fades within 24 h. Indeed, our results indicate the opposite: that graph ratings were slightly lower during the second session. These results suggest that while an educational intervention did not eliminate the truncation effect, its influence was durable for at least a day.

Study 5

In Studies 1–4, there were not statistically significant relationships between participant graph literacy and the size of the truncation effect. This measure is well-regarded, measuring relatively basic knowledge about graphs (Okan, Galesic, & Garcia-Retamero, 2016). However, while graph literacy is useful in certain contexts, it may not reflect the deeper expertise that develops over many years. Empirical evidence on the potential influence of expertise on misleading graphs is sparse, as individual variation in expertise is infrequently reported or manipulated. For example, Taylor and Anderson (1986) reported the results of a mini-experiment (7 pairs of graphs) administered to loan officers, but did not include any methodological details (e.g., sample size, means) or compare to non-experts. Another study investigated whether experts' interpretations of bar versus line graphs would be influenced by format (bar versus line); however, this work does not report a within-study comparison to non-experts (Peebles & Ali, 2015) and does not speak to distorted graphs.

In Study 5, we examined the size of the truncation effect in two doctoral student populations: PhD students pursuing quantitative fields versus the humanities. We reasoned that these samples would be comparable demographically but differ in their expertise with data visualization, providing a useful contrast between groups that are both highly educated but likely differ in experience with graphs. We expected that PhD students in quantitative fields would be unlikely to exhibit the truncation effect, which is the result of a relatively straightforward axis manipulation, particularly after they are given an explicit warning about the nature of the distortion.

Method

Participants. We recruited 165 PhD students (81 women, 2 non-binary gender; M_{age} = 26.87 years, SD_{age} = 3.58) pursuing degrees in quantitatively-oriented fields and 165 PhD students (109 women, 5 non-binary gender, M_{age} = 28.98, SD_{age} = 4.47) pursuing degrees in the humanities. Using publicly available email addresses and after consulting with the universities' respective institutional review boards, we recruited PhD students from programs in statistics, psychology, and economics, as well as in history and English, for our two samples. We aimed to recruit at least as many participants as in Study 4, especially anticipating a smaller truncation effect size than in the previously sampled populations. Pragmatically, given the difficulty of recruiting this specialized population, we aimed to recruit as many participants as possible in a reasonable time frame. Results for each sample were not analyzed until data collection was complete.

Reports of formal statistical training and graph literacy scale scores support our assumptions about statistical and data visualization experience within the two populations. Most quantitative participants (92%) reported having statistical training, whereas only 7% of humanities participants reported the same. Similarly, the quantitative sample scored higher on the graph literacy measure on average (M = 45.79, $SD_{\text{graph literacy}}$ = 7.18) than participants from the humanities (M = 37.59, $SD_{\text{graph literacy}}$ = 8.42). This difference in self-reported graph literacy was statistically significant: $t(327) = 9.52$, $p < .0001$, 95% CI of the difference [6.51, 9.91], Cohen's d = 1.05 [0.82, 1.28].

Participants were given the option to enter a raffle for $100 USD Amazon gift certificates; they were not otherwise compensated. No participants were excluded from analyses.

Materials. We used the same set of 40 bar graphs used in Study 1. As before, we assessed graph literacy using a 10-item scale.

Procedure. Participants followed the same procedures described for Study 3: axis truncation was manipulated within-subject, while warning was manipulated between-subjects. The structure and content of the warning was identical those given in the previous studies. Participants were successful when answering the instructional item, with 97.5% and 97.7% of quantitative and humanities participants answering the question correctly prior to feedback, respectively.

As in previous studies, participants completed the graph literacy assessment, demographic questions, and debriefing questions about graphs at the end of the study. We also asked participants questions about their formal and informal levels of statistical training.

Results

The truncation effect. We first replicated the truncation effect: the average graph rating was higher for truncated graphs than for control graphs: M_{control} = 3.73, SD = 0.54; $M_{\text{truncated}}$ = 4.17, SD = 0.76. This main effect was statistically significant: $t(329) = 13.9$, $p < .0001$, 95% CI of the difference [0.37, 0.50], Cohen's d = 0.48 [0.32, 0.63]. Most participants from both fields showed an overall truncation effect: 78.22% and 75.22% respectively, for humanities and quantitative students.

Next, we computed a linear mixed-effects model with graph type (0 = *control*, 1 = *truncated*), warning condition (0 = *no warning*, 1 = *warning*), and field (0 = *humanities*, 1 = *quantitative*) as binary fixed factors. The outcome variable was graph ratings; participant and item were modeled as random effects. We found statistically significant effects of graph type, warning condition, and field (Table 3). We also found statistically significant interactions between graph type and warning condition, graph type and field, and field and warning condition. However, the 3-way interaction between field, warning condition, and graph type was not statistically significant (p = .13). We ran follow-up models to clarify these interactions, described below. To preview, Figure 6 summarizes our results: we replicated the finding that an explanatory warning reduces but does not eliminate the truncation effect. We also found a smaller truncation effect in participants from quantitative fields when no warning was given, compared to participants from the humanities. However, this field advantage was not evidenced in the warning condition.

Graph type and warning condition. First, we investigated the interaction between graph type and warning condition to replicate results from Study 3 and 4, collapsing across field.

Table 3
Results for Study 5 from a Linear Mixed Effects Model Predicting Graph Ratings.

	Estimate (*b*)	95% CI of the Estimate	*SE*	|*t*|	*p*
Intercept	3.90	[3.56, 4.35]	0.21	18.22	< .0001
Graph Type (Control or Truncated)	0.69	[0.63, 0.75]	0.03	20.30	< .0001
Warning Condition (No Warning or Warning)	−0.30	[−0.47, −0.11]	0.09	3.31	.001
Field (Humanities or Quantitative)	−0.20	[−0.37, −0.04]	0.09	2.19	.03
Graph Type × Warning Condition	0.42	[−0.51, −0.32]	0.05	8.59	< .0001
Field × Graph Type	−0.14	[−0.23, −0.03]	0.05	2.97	.003
Field × Warning Condition	0.32	[0.06, 0.57]	0.13	2.47	.01
Field × Graph Type × Warning Condition	0.10	[−0.04, 0.24]	0.07	1.50	.13

Note. Experimental conditions (graph type, warning or no warning, humanities or quantitative) were dummy-coded. No Warning and Humanities served as the reference groups for the between-subjects manipulations. Control graphs served as the reference group for the within-subject manipulation.

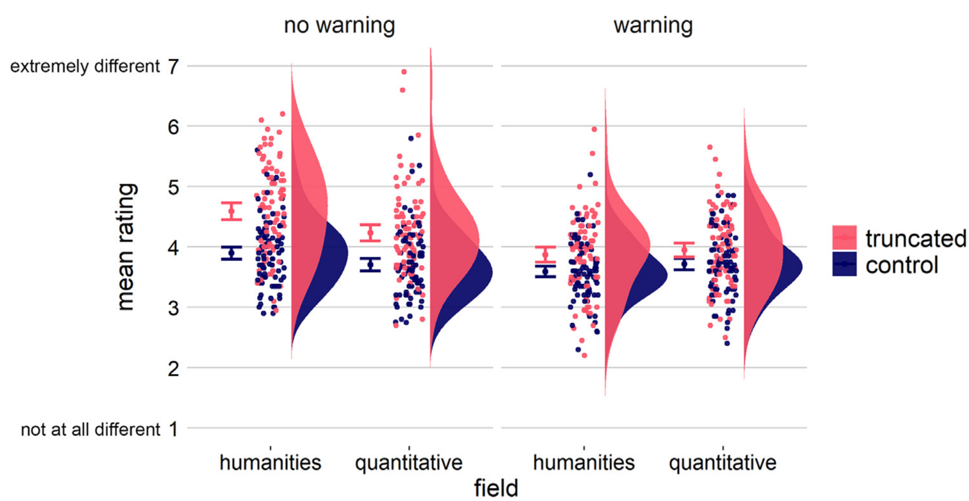

Figure 6. Raincloud plot for Study 5. Error bars reflect correlation- and difference-adjusted 95% confidence intervals of the means. Points represent each participant twice. The truncated versus control graphs variable was manipulated within-subjects.

We found that in both no warning (*estimate of the difference* = 0.62, 95% CI [0.57, 0.67], *SE* = 0.02, $p < .0001$) and warning ($estimate_{\mathrm{diff}}$ = 0.25, 95% CI [0.20, 0.30], *SE* = 0.02, $p < .0001$) conditions, ratings for truncated graphs were higher than control graphs. Next, we computed the effect of an explanatory warning for both control and truncated graphs. We found that an explanatory warning lowered graph ratings for both control ($estimate_{\mathrm{diff}}$ = 0.14, 95% CI [0.02, 0.27], *SE* = 0.06, $p = .03$) and truncated graphs ($estimate_{\mathrm{diff}}$ = 0.51, 95% CI [0.38, 0.64], *SE* = 0.06, $p < .0001$). The contrast between these two differences (i.e., warning's effect on control graphs compared to warning's effect on truncated graphs) was statistically significant ($p < .0001$), such that warnings lowered ratings more for truncated graphs than control graphs ($estimate_{\mathrm{diff}}$ = 0.37, 95% CI [0.30, 0.44], *SE* = 0.03). Thus, an explanatory warning reduced the size of the truncation effect (as in previous studies); it may have also led participants to be slightly more cautious in rating the differences shown by all graph types, not just truncated graphs.

Field and graph type. To explore the interaction between field and graph type, we computed contrasts of field for each level of graph type (control and truncated), averaging across levels of warning condition. We found that participants ratings for control graphs did not vary by field: 95% CI estimate of the difference [-0.09, 0.17], $p = .54$). However, the contrast between humanities and quantitative field participants for truncated graphs was statistically significant: $estimate_{\mathrm{diff}}$ = 0.13, 95% CI [0.01, 0.26], *SE* = 0.06, $p = .04$. Comparing these differences statistically, we found that the truncation effect (the difference between truncated and control graphs) was smaller for quantitative participants compared to humanities participants ($estimate_{\mathrm{diff}}$ = 0.09, 95% CI [0.02, 0.16], *SE* = 0.03, $p = .007$). This is consistent with the hypothesis that many years of quantitative expertise would translate into reduced susceptibility to the truncation effect.

Field and warning condition. To explore the interaction between field and warning condition, we computed contrasts of field for each level of the warning condition, averaging across levels of graph type. When no warning was given, we found a statistically significant effect of field, such that participants from the humanities rated graphs higher than participants from quantitative fields: $estimate_{\mathrm{diff}}$ = 0.27, 95% CI [0.10, 0.44], *SE* = 0.09, $p = .002$. However, when a warning was provided, the effect of field was not statistically significant: 95% CI of the difference [−0.27, 0.07], $p = .26$. Then, we statistically compared the differences by field (humanities – quantitative) between the no warning and warning conditions. We found that this difference was statistically significant ($estimate_{\mathrm{diff}}$ = 0.37, 95% CI

Please cite this article in press as: Yang, B. W., et al. Truncating Bar Graphs Persistently Misleads Viewers. *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2020.10.002

[0.13, 0.62], SE = 0.12, p = .003). In other words, when not given a warning, humanities participants rated graphs higher than participants from quantitative fields.

Overall, we found that the truncation effect was persistent across all conditions. PhD students from quantitative fields exhibited a smaller truncation effect compared to PhD students from the humanities when no warning was given. However, when given a warning, we found that the participants from the humanities "caught up" with those from quantitative fields. We note that even PhD students in quantitative fields given an explanatory warning about the nature of the manipulation immediately preceding graph ratings showed a truncation effect, *estimate* = 0.24, 95% CI [0.14, 0.34], $t(82) = 4.72$, $p < .0001$. So, while educational training and explicit warnings may reduce the size of the truncation effect, they do not eliminate it.

General Discussion

Across five studies, we empirically investigated the *truncation effect*, showing that quantity differences were judged as larger in truncated compared to control graphs. This effect was reliable (Cohen's d = 0.48–1.26) and observed across most participants (83.5% across studies), regardless of graph literacy scores. Teaching people about how truncation can mislead attenuated the truncation effect somewhat, both immediately and after a 1-day delay, but failed to eliminate the effect. Indeed, the truncation effect persisted across the five studies in the present work, contexts which likely engendered greater than typical vigilance for graph distortions. Namely, participants made judgments in an experiment labeled as having to do with graphs; they then viewed 40 graphs in a within-subject design allowing for frequent comparisons of control and distorted graphs. Participants were sometimes given a clear explanation about the single form of distortion they were immediately about to encounter (Studies 2–5), and in Study 5, participants were pursuing doctoral degrees, half of whom were pursuing studies within quantitative fields. Our materials were neutral in nature, and unlikely to encourage motivated reasoning processes (Kunda, 1990). These conditions are in stark contrast to a user casually consuming news via social media on a mobile device, in a context where misinformation and graph distortions can take a myriad of forms, rarely come with a clear warning, and may interact with existing beliefs. Our work fills a much needed methodological and theoretical gap in the literature, bringing together the strengths of behavioral science methods to inform long-standing questions around graph distortion.

Effects of Graph Literacy

Why did we not see effects of graph literacy in predicting the size of the truncation effect? This may seem initially contradictory, given that participants with quantitative training in Study 5 showed reduced truncation effects in the absence of a given warning. Our results suggest a distinction between the construct of subjective graph literacy (captured by items such as "How often do you find graphical information to be useful?") and the deeper expertise developed over years in doctoral programs in statistics, psychology, and economics. We speculate that graph literacy relates to prior knowledge related to attending to and processing "arbitrary graph conventions" such as axis labels and graph titles, reflecting basic training in understanding graphs (Okan et al., 2016). Thus, graph literacy may be more predictive in contexts when accurate interpretation relies on such graphical training, such as with more complex data visualizations. However, comparing the height of two bars is intuitive, reflective of mappings corresponding to experiences in the physical world (Shah & Hoeffner, 2002; Tversky, 2011). Our simpler graph stimuli are representative of graphs in mass media, where most data visualizations are designed to be meaningful and communicative to a general audience.

Possible mechanisms. The present work establishes the truncation effect in bar graphs and the extent to which this effect is persistent. Our attempts to mitigate truncation effects by directly warning participants about y-axis truncation were mixed, suggesting that the problem was not just about a lack of knowledge. Future experimental work should be designed to identify explanations behind these observations. Specifically, one possibility is that the appearance of a large difference initially anchors judgments, consistent with previous research showing that initially presented values have disproportionate influence (Furnham & Boo, 2011; Tversky & Kahneman, 1974). Although most of the literature has focused on numeric anchors, recent articles examined the extent to which this phenomenon generalizes to visual stimuli (Langeborg & Eriksson, 2016; LeBoeuf & Shafir, 2009). In previous work, warnings were much less effective in the case of visually driven phenomena such as change blindness (Simons, 2000) or optical illusions (e.g., Barlow & Hill, 1963; Williams & Yampolskiy, 2018); visual anchoring as a phenomenon could be compared and contrasted with these phenomena.

Another and non-exclusive explanation relates to the norm of communication to trust the speaker, or this case, the graph maker. Listeners tend to assume that speakers are providing truthful, relevant, and clear information (Grice, 1975). Thus, the reader of a graph may assume that the difference highlighted in a graph represents a meaningful unit of comparison. Future empirical studies around how people perceive, make judgments about, and later remember graphs can contribute to expanding existing theories of graph comprehension (Carpenter & Shah, 1998; Pinker, 1990; Shah, Freedman, & Vekiri, 2005).

To Truncate or Not to Truncate

Data visualization experts have recommended that the y-axes of bar graphs specifically should start at zero (e.g., "If your numbers are represented by the length or height of objects—bars, in this case—the length of height should be *proportional* to those numbers"; Cairo, 2019). Our data support this recommendation empirically. However, our recommendation to not truncate vertical axes is specific to bar graphs. Line graphs and dot plots, for example, do not represent numerical values as continuous visual areas and truncation may be appropriate in these cases (Cleveland, 1994; Zoss, 2016, p. 42). In other words, if length or height is not the primary means of communicating numerical quantities, a zero baseline may not be necessary. When small numerical differences matter—such as a single degree shift in average global temperature—there are better formats than bar graphs. We suggest future research make clear the differences between bar and line graphs, and their distinct affordances in illustrating differences.

Implications and Future Directions

In light of the present work, we suggest that the burden for accurate presentation of data in bar graphs falls on the shoulders of *graph makers*. Our sophisticated sample of graph consumers in Study 5 were still susceptible to the truncation effect, even though they were given a warning and placed in a context likely to engender skepticism (a scientific study about graphs). Viewing graphs as arguments, these results suggest that the practice of truncating the y-axis of a bar graph is comparable to a particularly persuasive rhetorical fallacy, and thus, should be avoided.

Given the ubiquity of bar graphs and the relative ease of creating them, this work has important implications for the clear and responsible communication of data. We urge producers of graphs (including many of our present readers) to avoid the practice of truncating y-axes of bar graphs, contributing to stronger cultural norms for responsible data visualization. Our recommendation is consistent with more general advice given by risk communication specialists: while communicators would prefer that people process information systematically, consumers are much more likely to digest information heuristically, and thus are extremely susceptible to variations in presentation format (Visschers, Meertens, Passchier, & Vries, 2009). Ideally, data visualizations should be crafted such that even initial impressions are well-aligned to the conclusions afforded by more careful analysis of underlying numerical trends.

However, we acknowledge that it is unrealistic to expect the solution to this problem to be the removal of all misleading graphs. Our work demonstrates that an explanatory warning can reduce the size of the truncation affect and that this protective effect lasts for at least a day, suggesting an intervention approach that is helpful, if incomplete. We hope that the present work serves as a catalyst for future empirical work exploring the impact of data visualizations. We suggest that future studies may explore other common graph types (e.g., line graphs), the memorability of types of graphs, continued interventions, and testing predictions from different theoretical frameworks for explaining the truncation effect.

Author Contributions

C. Vargas Restrepo, B. W. Yang, and E. J. Marsh developed the study concepts and design. Testing and data collection were performed by C. Vargas Restrepo and B. W. Yang. Statistical analyses were conducted by C. Vargas Restrepo and B. W. Yang, with assistance and critical feedback from M. L. Stanley. C. Vargas Restrepo and B. W. Yang drafted the manuscript, and M. L. Stanley and E. J. Marsh provided critical revisions. All authors approved the final version of the manuscript for submission.

Open Practices Statement

Experiments reported in this article were not formally preregistered. All stimuli, de-identified data, and reproducible analyses can be accessed at osf.io/ytq3h. We have done our best to ensure that analyses can be successfully reproduced. We are happy to answer questions and troubleshoot. Please contact the corresponding author at by34@duke.edu.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the James S. McDonnell Foundation and 2 Undergraduate Research Grants from Duke University. B. W. Yang is supported by an NSF Graduate Research Fellowship. We thank Kevin O'Neill for providing consultations regarding data analysis and the Marsh Lab, Priti Shah, Morton Ann Gernsbacher, Steve Lindsay, and Henry L. Roediger, III for helpful comments on previous versions of this manuscript.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. (2018). Raincloud plots: A multi-platform tool for robust data visualization. *PeerJ Preprints*, https://doi.org/10.7287/peerj.preprints.27137v1

Barlow, H. B., & Hill, R. M. (1963). Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects. *Nature*, *200*(4913), 1345–1347. https://doi.org/10.1038/2001345a0

Beattie, V., & Jones, M. (1999). *Financial graphs: True and fair?* Australian CPA.

Beattie, V., & Jones, M. J. (1992). The use and abuse of graphs in annual reports: Theoretical framework and empirical study. *Accounting and Business Research; Abingdon*, *22*(88), 291.

Beattie, V., & Jones, M. J. (2008). Corporate reporting using graphs: A review and synthesis. *Journal of Accounting Literature*, *27*, 71–110.

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

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., et al. (2013). What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2306–2315. https://doi.org/10.1109/TVCG.2013.234

Cairo, A. (2019). *How charts lie: Getting smarter about visual information*. W. W. Norton & Company.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology Applied*, *4*(2), 75–100. https://doi.org/10.1037/1076-898X.4.2.75

Cho, C. H., Michelon, G., & Patten, D. M. (2012). Impression management in sustainability reports: An empirical investigation of the use of graphs. *Accounting and the Public Interest*, *12*(1), 16–37. https://doi.org/10.2308/apin-10249

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., et al. (2019). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, https://doi.org/10.1007/s11109-019-09533-0

Cleveland, W. S. (1994). *The elements of graphing data (Rev. ed)*. AT&T Bell Laboratories.

Cleveland, W. S., Diaconis, P., & Mcgill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. *Science*, *216*(4550), 1138–1141. https://doi.org/10.1126/science.216.4550.1138

Courtis, J. K. (1997). Corporate Annual Report Graphical Communication in Hong Kong: Effective or Misleading? *Journal of Business Communication*, *34*(3), 269–284. https://doi.org/10.1177/002194369703400304

Cousineau, D. (2017). Varieties of confidence intervals. *Advances in Cognitive Psychology*, *13*(2), 140. https://doi.org/10.5709/acp-0214-z

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.

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087–1100. https://doi.org/10.3758/MC.38.8.1087

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten*. Analytics Press.

Fischer, C. (2008). Feedback on household electricity consumption: a tool for saving energy? *Energy Efficiency*, *1*(1), 79–104. https://doi.org/10.1007/s12053-008-9009-7

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-economics*, *40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy: A cross-cultural comparison. *Medical Decision Making*, *31*(3), 444–457. https://doi.org/10.1177/0272989X10373805

Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: Overcoming challenges in people's understanding of risks. *Social Science & Medicine*, *70*(7), 1019–1025. https://doi.org/10.1016/j.socscimed.2009.11.031

Garcia-Retamero, R., Cokely, E. T., Ghazal, S., & Joeris, A. (2016). Measuring graph literacy without a test: A brief subjective assessment. *Medical Decision Making*, *36*(7), 854–867. https://doi.org/10.1177/0272989X16655334

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *2*(1), 231–239.

Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, *47*(2), 183–210. https://doi.org/10.1080/03057267.2011.605307

Greene, E., Flynn, M. S., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 207–219. https://doi.org/10.1016/S0022-5371(82)90571-0

Gregory, R. L. (2009). *Seeing through illusions*. Oxford University Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics, vol. 3: Speech acts* (Vol. 3) (pp. 41–58). Academic Press.

Huff, D. (1954). *How to lie with statistics*. W. W. Norton & Company.

Jensen, O. M. (2003). *Visualisation turns down energy demand (Conference) | ETDEWEB*. pp. 1–3. https://www.osti.gov/etdeweb/biblio/20501271

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Langeborg, L., & Eriksson, M. (2016). Anchoring in numeric judgments of visual stimuli. *Frontiers in Psychology*, *7* https://doi.org/10.3389/fpsyg.2016.00225

LeBoeuf, R. A., & Shafir, E. (2009). Anchoring on the "here" and "now" in time and distance judgments. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *35*(1), 81–93. https://doi.org/10.1037/a0013665

Lenth, R. (2019). *Emmeans: Estimated Marginal Means, aka Least-Squares Means (Version R package version 1.3.2) [Computer software]*. https://CRAN.R-project.org/package=emmeans

Leo, S. (2019). *Mistakes, we've drawn a few*. Medium. https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368

Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, *34*(5), 1140–1149.

Mather, D., Mather, P., & Ramsay, A. (2005). An investigation into the measurement of graph distortion in financial reports. *Accounting and Business Research*, *35*(2), 147–160. https://doi.org/10.1080/00014788.2005.9729670

Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G., & Garry, M. (2013). On the (non)persuasive power of a brain image. *Psychonomic Bulletin & Review*, *20*(4), 720–725. https://doi.org/10.3758/s13423-013-0391-6

Mogull, S. A., & Stanfield, C. T. (2015). Current use of visuals in scientific communication. In *2015 IEEE international professional communication conference (IPCC)*. pp. 1–6. https://doi.org/10.1109/IPCC.2015.7235818

Newman, E. J., Garry, M., Bernstein, D. M., Kantner, J., & Lindsay, D. S. (2012). Nonprobative photographs (or words) inflate truthiness. *Psychonomic Bulletin & Review*, *19*(5), 969–974. https://doi.org/10.3758/s13423-012-0292-0

Okan, Y., Galesic, M., & Garcia-Retamero, R. (2016). How people with low and high graph literacy process health graphs: Evidence from eye-tracking: Graph literacy and health graph processing. *Journal of Behavioral Decision Making*, *29*(2–3), 271–294. https://doi.org/10.1002/bdm.1891

Okan, Y., Garcia-Retamero, R., Galesic, M., & Cokely, E. T. (2012). When higher bars are not larger quantities: On individual differences in the use of spatial information in graph comprehension. *Spatial Cognition and Computation*, *12*(2–3), 195–218. https://doi.org/10.1080/13875868.2012.659302

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. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. pp. 1469–1478. http://dl.acm.org/citation.cfm?id=2702608

Peebles, D., & Ali, N. (2015). Expert interpretation of bar and line graphs: The role of graphicacy in reducing the effect of graph format. *Frontiers in Psychology*, *6* https://doi.org/10.3389/fpsyg.2015.01673

Penrose, J. M. (2008). Annual report graphic use: A review of the literature. *The Journal of Business Communication (1973)*, *45*(2), 158–180. https://doi.org/10.1177/0021943607313990

Peterson, B. K. (1983). Tables and graphs improve reader performance and reader reaction. *Journal of Business Communication*, *20*(2), 47–55. https://doi.org/10.1177/002194368302000204

Pinker, S. (1990). A theory of graph comprehension. In *In Artificial intelligence and the future of testing*. pp. 73–126.

Practices, N. G. A. C. B., & Council of Chief State School Officers. (2010). *Common core state standards (Mathematics)*. http://www.corestandards.org/Math/Content/1/MD/

Raschke, R. L., & Steinbart, P. J. (2008). Mitigating the effects of misleading graphs on decisions by educating users about the principles of graph design. *Journal of Information Systems*, *22*(2), 23–52. https://doi.org/10.2308/jis.2008.22.2.23

Roberts, S. (2020). *Flattening the coronavirus curve*. The New York Times. https://www.nytimes.com/2020/03/11/science/coronavirus-curve-mitigation-infection.html

Roser, M., Ritchie, H., & Ortiz-Ospina, E. (2020). *Coronavirus disease (COVID-19) – Statistics and research. Our world in data*. https://ourworldindata.org/coronavirus

Sargent, S. L. (2007). Image effects on selective exposure to computer-mediated news stories. *Computers in Human Behavior*, *23*(1), 705–726. https://doi.org/10.1016/j.chb.2004.11.005

Shah, P., & Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes: Topics in cognitive science. *Topics in Cognitive Science*, *3*(3), 560–578. https://doi.org/10.1111/j.1756-8765.2009.01066.x

Shah, P., & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*(1), 47–69.

Shah, P., Freedman, E., & Vekiri, I. (2005). The comprehension of quantitative information in graphical displays. In *In the cambridge handbook of visuospatial thinking* (1st ed., pp. 426–477). Cambridge University Press.

Shere, D. (2012). *Dishonest fox chart: Bush tax cut edition*. Media Matters for America. https://www.mediamatters.org/fox-business/dishonest-fox-chart-bush-tax-cut-edition

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, *7*(1–3), 1–15. https://doi.org/10.1080/135062800394658

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*(2), 73–74.

Steinbart, P. J. (1989). *The auditor's responsibility for the accuracy of graphs*. https://search.proquest.com/openview/a3116ecb8f7a561436180910fa3a5da0/1?pq-origsite=gscholar&cbl=3330

Taylor, B., & Anderson, L. (1986). Misleading graphs guidelines for the accountant. *Journal of Economics Business and Accountancy Ventura*, *162*(4), 126–135.

Tufte, E. R. (2001). *The visual display of quantitative information* (2. ed). Graphics Press.

Tversky, B. (2011). Visualizing thought. *Topics in Cognitive Science*, *3*(3), 499–535. https://doi.org/10.1111/j.1756-8765.2010.01113.x

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Tversky, B., & Schiano, D. J. (1989). Perceptual and conceptual factors in distortions in memory for graphs and maps. *Journal of Experimental Psychology General*, *118*(4), 387.

Visschers, V. H. M., Meertens, R. M., Passchier, W. W. F., & Vries, N. N. K. D. (2009). Probability information in risk communication: A review of the research literature. *Risk Analysis*, *29*(2), 267–287. https://doi.org/10.1111/j.1539-6924.2008.01137.x

Wilkinson, L. (2006). *The grammar of graphics*. Springer Science & Business Media.

Williams, R. M., & Yampolskiy, R. V. (2018). *Optical illusions images dataset. ArXiv:1810.00415 [Cs]*. http://arxiv.org/abs/1810.00415

Witt, J. K. (2019). Graph construction: An empirical investigation on setting the range of the Y-Axis. *Meta-Psychology*, *3*.

Zacks, J., Levy, E., Tversky, B., & Schiano, D. J. (1998). Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology Applied*, *4*(2), 119–138. https://doi.org/10.1037/1076-898X.4.2.119

Zoss, A. (2016). Designing public visualizations of library data. In L. Magnuson (Ed.), *Data visualization: A Guide to visual storytelling for libraries*. Rowman & Littlefield.

Received 30 May 2020;
received in revised form 14 October 2020;
accepted 14 October 2020
Available online xxx