An Investigation Into the Cognitive Processing of Magnitude in Data Visualisations

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2023

Duncan Bradley

School of Health Sciences

Table of Contents

Word Count:

List of Figures

List of Tables

Abstract

Data visualisations are effective communication tools which leverage the power of the human visual and cognitive systems. However, data visualisation design choices can substantially influence interpretations of presented information. Interpreting data visualisations does not just involve accurately comprehending values, but often also involves drawing inferences about data and making subjective judgements. Therefore, developing an understanding of cognitive processing of data visualisations is important for guiding the construction of effective and faithful representations of data.

Data visualisations can convey many different features of a dataset. One such feature is the absolute magnitude of plotted values: how large or small they are. This thesis investigates cognitive mechanisms involved in judging absolute magnitude in data visualisations, revealing the effects of design choices on interpretation.

The first set of experiments in this thesis (three experiments) explores the role of axis limits in informing magnitude judgements. Manipulating axis limits in dot plots causes the same data points to appear near the top or bottom of the visualisation. Participants' responses revealed an association between higher positions and higher magnitude ratings, indicating a bias in interpretation. Two further experiments employing dot plots with inverted y-axes indicated that impressions of magnitude were driven primarily by the relative positions of data points within axis limits, not their absolute physical positions.

The second experiment in this thesis extends inquiry into the role of axis limits to choropleth maps. Manipulating the limits of accompanying colour legends alters the framing of presented data, without changing how plotted values appeared. Extending the colour legend's upper limit beyond the maximum plotted value resulted in lower magnitude ratings. This demonstrates that interpretations of absolute magnitude are informed by surrounding context, not just by the appearance of plotted values.

The final set of experiments in this thesis (two experiments) explores how additional knowledge about plotted data can inform interpretations of magnitude. Denominators provide numerical context relevant to magnitude judgements. Extending a bar chart's axis beyond plotted data to incorporate a denominator value elicited lower magnitude ratings, compared to bar charts' default settings. Omitting denominator information from accompanying text substantially increased this bias. This illustrates that additional knowledge about a dataset diminishes the roles of axis limits in informing impressions of magnitude.

This work was conducted with a focus on computational reproducibility. In addition to sharing data and analysis code, this involved facilitating the ability to reproduce the computational environment used for analysis. This approach, which increases openness and transparency in research, is also discussed in detail.

Through experimental research, this thesis reveals how the framing of values within axes informs judgements of their absolute magnitudes. The results provide insight into the cognitive processing of magnitude in data visualisations, wherein context shapes viewers' inferences. This illustrates how inevitable subjectivity in data visualisation design can influence a data visualisation's appearance and its message. Data visualisation designers should consider the graphical representation of absolute magnitude and, where appropriate, employ axes ranges which faithfully convey this aspect of data.

Declaration

- One experiment reported in Chapter 4 was conducted in collaboration with Gabriel Strain, who submitted a report on this research written independently in partial fulfilment of his MRes Psychology degree at the University of Manchester (2021). The report written within this thesis is my own work.
- The experiment reported in Chapter 5 was conducted in collaboration with Boshuo Zhang, who submitted a report on this research written independently in partial fulfilment of his MSc Advanced Computer Science degree at the University of Manchester (2022). The report written within this thesis is my own work.
- One experiment reported in Chapter 6 was conducted in collaboration with Harvey Schneider, who submitted a report on this research written independently in partial fulfilment of his MRes Psychology degree at the University of Manchester (2022). The report written within this thesis is my own work.
- No other portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

I would like to thank my supervisors, Andrew Stewart and Caroline Jay. Andrew's mentorship during my BSc and MRes degrees was foundational in motivating my PhD, and he has generously offered his time, expertise, and his fresh outlook, throughout the entire process. Caroline's perspective has also been extremely valuable and her advice and wisdom has been instrumental in improving my work. I am very grateful to have been given the opportunity to collaborate with such a dedicated team. I am also indebted to Jen McBride for the considerate and pragmatic guidance which she offered as my advisor.

For productive conversations, practical support, and camaraderie, thanks to Libby Le-Iuan, Gabe Strain, and Paul Stott. Thanks also to all in room 124e, and to Tristan Johnston-Wood and Sam Langton, for sharing experiences of the PhD process. I am grateful to Boshuo Zhang, Harvey Schneider, and Gabe Strain for their hard work in helping to develop experimental stimuli during their Master's programmes.

Chapter 1

Introduction

Data visualisations help people make sense of numbers. Whilst a list of numbers may allude to an upwards trend, presenting those numbers *as* visual phenomena can aid interpretation. A data visualisation can systematically depict the precise numerical values, facilitating mental processing of this information.

As *external* representations of data, visualisations reduce perceptual and cognitive burdens in interpretation (Scaife and Rogers, 1996). By imparting efficiency and clarity, data visualisations support pattern-recognition and reasoning. However, a single dataset can be depicted in numerous ways, and different designs can vary widely in their effectiveness (Franconeri et al., 2021). Thus, data visualisation's strength can also be its vulnerability. Outsourcing cognitive processes to a graphical depiction leaves a viewer at the mercy of the chosen method of visual representation. Thus, understanding successful design is crucial.

The effectiveness of data visualisations can be defined in many ways, encompassing their various objectives, which include informing, persuading, engaging, and promoting retention of information (Bertini et al., 2020). However, in general, successful data visualisations will convey pertinent information in a visually- and cognitively-comprehensible manner (Mackinlay, 1986; Wijk, 2005). Failing to meet these criteria risks misleading viewers, which is antithetical to the purpose of data visualisation. Therefore, knowledge of human factors in visualisation is vital for ensuring charts, graphs, and maps achieve their potential.

1.1 Research Motivation and Objective

Interpreting data in any medium does not involve simply observing numerical values, but rather identifying patterns and making *inferences* about the data. For example, one may notice that values are decreasing rapidly, or that there is substantial variability in the dataset, or that some data are missing. One may also make inferences about how large or small values are. This is an important aspect of understanding data, because the same numerical value can be considered large or small depending on its context.

The BBC radio programme *More or Less*, which examines statistics reported in the news and elsewhere, often addresses this issue. Some figures may instinctively seem large, others small. However, asking the question 'Is it a big number?' considers whether this initial impression is appropriate. For example, a country's multi-trillion dollar national debt may appear high, but its level of debt may be similar to that of other countries when accounting for its high GDP (McDonald, 2023). Similarly, understanding the context of the Richter scale is required to determine that a value less than one does indeed reflect a small earthquake magnitude (Harford, 2022). Gauging magnitude is important for interpreting numbers in a wide variety of situations.

Because data visualisations are used to convey numerical information, it is important to understand how they may influence judgements of the magnitude of presented values. Studying the cognitive processing of magnitude can reveal how inferences are generated and provide insight into the effects of data visualisation design choices. This, in turn, can inform recommendations for designers who may wish to represent the magnitude of values using graphical cues. Yet, this has been an under-explored topic, with insufficient empirical research exploring this aspect of interpreting visualisations. The objective of this thesis is to generate robust empirical evidence on the interpretation of magnitude in data visualisations.

1.2 Contributions

Through a series of empirical experiments, I demonstrate that data visualisation design choices can affect mental representation of numbers' magnitudes. These large-sample, controlled experiments strengthen and expand the existing evidence base on this overlooked aspect of data visualisation. I reveal that, in various types of visualisations, graphical cues to context play a role in the processing of how large or small values are. Specifically, judgements about magnitude are informed by the relative positions of values within axis limits. Focusing on underlying cognitive mechanisms generates findings which are applicable to a variety of visualisation formats. These findings also contribute recommendations for designers, which involve considering suitable axis ranges in order to convey magnitude appropriately. This guidance challenges a convention in data visualisation design and advocates against the use of particular default settings, where appropriate.

1.3 Overview of the Thesis

In Chapter 2, I provide context for the empirical work conducted in this project, by reviewing related research. I advocate for an evidence-based approach to data visualisation design which takes into account the mental processing of information. I discuss perceptual and cognitive biases in the interpretation of visualisations, and with a particular focus on the role of numerical axes. This chapter justifies the focus of my investigation and explains how it is informed by prior work in this area.

In Chapter 3, I discuss the methodology and epistemological approach employed in this thesis. In addition to explaining my chosen experimental and statistical techniques, I describe how research projects can benefit from efforts to increase transparency. Approaches for increasing the reproducibility of research are presented, in line with recommendations from a variety of disciplines. A particular emphasis is placed on computational reproducibility: the capacity to recreate the computational environment used in generating results. This provides background for the practices used in the following empirical research chapters.

In Chapter 4, I present a set of three experiments which establish that interpretations of magnitude can be influenced by data visualisation designs. The first experiment demonstrates that manipulating axis limits in dot plots affects participants' judgements of overall magnitude. Two further experiments investigate whether this occurs because axis limits alter the absolute or relative positions of plotted values within axis limits. In dot plots with inverted y-axes, where higher numerical values are presented at lower positions, values near bottom were associated with higher magnitudes. This illustrates that interpretations of magnitude are informed by the relative positions of values within axis limits.

In Chapter 5, I present an experiment which explores how visualisations may influence interpretations of magnitude even when the appearance of plotted values remains unchanged. This experiment demonstrates that manipulating colour legend limits in choropleth maps affects participants' judgements of overall magnitude. Participants rated magnitudes as lower when the range of values on the colour legend extended beyond the largest plotted value. This illustrates that the numerical context accompanying plotted values can influence interpretations of magnitude, without altering the physical appearance of those values.

In Chapter 6, I present a set of two experiments which investigate the role of contextual information on interpretations of magnitude. The first experiment demonstrates that participants' judgements of overall magnitude were affected by extension of bar charts' upper axis limits which incorporated a denominator value. A second experiment reveals that participants' bias was increased when this denominator information was excluded from the text accompanying the chart. This illustrates that knowledge about a dataset's characteristics (e.g., denominator value) can influence the extent to which design choices affect interpretations of magnitude.

Finally, in Chapter 7, I present a synthesis of this empirical work, alongside a discussion of implications and future directions.

References

Bertini E, Correll M, Franconeri S. 2020. Why Shouldn't All Charts Be Scatter Plots? Beyond Precision-Driven Visualizations. *arXiv:200811310 [cs]*.

Franconeri SL, Padilla LM, Shah P, Zacks JM, Hullman J. 2021. The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest* **22**:110–161. doi:10.1177/15291006211051956

Harford T. 2022. Mini-budget, TV audience for the Queen's funeral and 0.5 on the Richter scale. *More or Less*.

Mackinlay J. 1986. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics* **5**:110–141. doi:10.1145/22949.22950

McDonald C. 2023. US national debt: Is $32 trillion a big number? *More or Less*.

Scaife M, Rogers Y. 1996. External cognition: How do graphical representations work? *International Journal of Human-Computer Studies* **45**:185–213.

doi:10.1006/ijhc.1996.0048

Wijk JJ van. 2005. The Value of Visualization 8.

Preface to Chapter 5

The empirical research conducted in this thesis thus far has demonstrated that interpretations of magnitude are influenced by the numerical context in which plotted values are presented. However, this thesis has not yet explored denominator values, which are another relevant aspect of context for magnitude judgements. Knowledge of a denominator is relevant to judging absolute frequencies. For example, when evaluating a business expense, awareness of the total available budget can provide perspective on the magnitude of the amount spent. Including denominator values in axes is another example of conveying relevant values that do not appear in plotted data. Therefore, the third and final investigation in this thesis explores how this type of contextual information informs interpretations of magnitude in bar charts.

This final set of experiments can also assist in understanding how the influence of visualisation design choices on a viewer's interpretation is affected by their awareness of a dataset's characteristics. The first set of experiments in this thesis observed a relatively small bias in the interpretation of probability values in familiar scenarios. The second investigation in this thesis observed a relatively large bias in the interpretation of ambiguous pollution measurements. The third and final set of experiments in this thesis contains a manipulation of participants' knowledge about plotted datasets. This provides insight into how knowledge of the numerical context behind plotted data can affect magnitude judgements, in addition to graphical cues.

Chapter 2

Axis Limits and Denominator Information Influence Absolute Magnitude Ratings in Bar Charts

2.1 Abstract

Consider a statistic corresponding to the number of public transport users in a particular town. Gauging whether this number is large or small requires awareness of the total population (the denominator). In data visualisations, an axis which extends beyond plotted values can act as a graphical cue to a denominator value, but default upper axis limits (e.g., in *ggplot2*) are typically based on the highest plotted value. In two experiments (combined $N = 350$), we explore the influence of default and extended axes on interpretations of magnitude in bar charts. We observe that values plotted using default axes were rated as higher, compared to values plotted using extended axes. Furthermore, the absence of denominator information in accompanying text amplifies the effect of axis limits on judgements. Whereas prior work has often focused on judgements of the *differences between values*, this work contributes to an understanding of how viewers interpret the magnitudes of *the values themselves*. We also discuss implications for effective design, which involve considering both axis limits and accompanying contextual information.

2.2 Introduction

The question 'Is it a big number?' is often raised on the BBC radio programme *More or Less* when probing eye-catching statistics. A figure of several million pounds may initially seem large, but may represent a small proportion of total government spending. Awareness of a denominator value can influence judgement of a number's magnitude. In data visualisation, this contextual information can be displayed by extending an axis to accommodate the denominator value. However, this approach is infrequently used, since typical default axis settings are based on plotted data only. In this study, we explore how these axis limits affect interpretations of how large or small plotted values are.

2.2.1 Overview

Across two experiments, we investigate the interpretation of magnitudes in bar charts. We plotted fictitious datasets, which contained multiple observations with the same denominator value. In the first experiment, we displayed charts with default axes which terminated just above plotted values, or axes which extended to the denominator value specified in accompanying text. Participants rated values' magnitudes as higher when default axes were used, compared to extended axes. In the second experiment, we manipulated the axis limits, as before, and also the presence of the denominator information in accompanying text. Variation in responses to the two different axis settings was greater when denominator information was not supplied. This indicates that this information influences the biasing effect of a chart's appearance.

2.2.2 Related Work

2.2.2.1 Biased Interpretation of Numerical Information

A wealth of research demonstrates biases in the interpretation of numbers. A survival rate elicits different judgements of a disease compared to its corresponding mortality rate (Tversky and Kahneman, 1981), and the fat content of a meat product elicits different judgements of the product compared to its corresponding lean content (Levin, 1987). The units used to express the same values (e.g., months vs. years) affect comparisons (Burson et al., 2009; Monga and Bagchi, 2012) and also interpretations of precision and accuracy (Zhang and Schwarz, 2012).

Various biases in magnitude judgements reveal the importance of accounting for numerical context. Base rate neglect describes difficulty acknowledging population-level characteristics when making judgements about a sample (Cosmides and Tooby, 1996). Format neglect describes a bias against incorporating set size information when judging percentage formats (top 20%) and numerical formats (top 10, (Sevilla et al., 2018)). Denominator neglect involves over-weighting numerator information at the expense of denominator information (Reyna and Brainerd, 2008). The latter also leads (in part) to large percentages of a small number appearing greater than the *numerically equivalent* smaller percentages of a larger number (Li and Chapman, 2013). The general mechanism responsible for these biases is a failure to properly acknowledge numerical context.

2.2.2.2 Visualising Denominator Information

Data visualisations can help combat biases. Denominator information becomes visually available when icon arrays present both focal outcomes (e.g., number deceased) *and also* alternative outcomes (e.g., number survived). Research suggests that providing this visual cue to a denominator (e.g., total number at risk) facilitates reasoning. For example, including alternative outcomes in an icon array reduces denominator neglect, increasing comprehension of relative risk (Garcia-Retamero and Galesic, 2010). Icon arrays displaying both types of outcome are particularly helpful for understanding datasets with unequal denominators, and for individuals with high graph literacy (Okan et al., 2012). However, these effects are largest when depicting small probabilities (Okan et al., 2020).

Stacked bar charts function similarly to icon arrays: one bar represents the focal outcome, another bar represents the alternative outcome, and their combination represents the denominator. Like icon arrays, stacked bar charts diminish the influence of denominator neglect (Stone et al., 2003). Despite this, denominator information can be displayed in bar charts *without* using additional stacked bars to represent alternative outcomes. Extending an axis to incorporate a denominator value also communicates relevant numerical context. In this case, the blank space between the bars for the focal outcome and the upper axis limit corresponds to the alternative outcome. Research has demonstrated that bar charts representing alternative outcomes using blank space increase perceptions of risk likelihood compared to those representing alternative outcomes using stacked bars (Stone et al., 2017). However, this research did not manipulate the presentation of denominatorinformation, since identical axis limits were employed across conditions.

Directly manipulating bar charts' upper axis limits provides insight into the use of this cue to denominator information. Bar charts with axes which extend to the denominator value produce more accurate estimates of changes in risk (Garcia-Retamero and Galesic, 2010). This design also elicits decreased ratings of risk perception (Okan et al., 2018), though numerical labels reduce this effect.

Various accounts have sought to explain the effects of including denominator information in visualisations. Depicting denominators may facilitate understanding of part-to-whole relationships, diminishing the class-inclusion errors associated with denominator neglect (Reyna, 2008). This argument is consistent with Fuzzy Trace Theory, which also predicts the influence of physical attributes in gist representations (i.e, short bars may elicit smaller magnitude judgements, Reyna, 2008). Additionally, Stone et al. (2018) suggest that facilitation of *proportional reasoning* may be largely responsible for observed effects. They observed that increasing the salience of the denominator in *text* fails to affect judgements, yet a *graphical* representation effectively communicates the true scale of the denominator, helping put numerators into perspective.

2.2.2.3 Judgements of Absolute Magnitude

As discussed above, Garcia-Retamero and Galesic (2010) and Okan et al. (2018) demonstrate that extending axes to incorporate denominator values influences interpretation of risk. However, these studies do not provide specific evidence on how this affects interpretations of the *absolute magnitude* of plotted values.

Garcia-Retamero and Galesic (2010) measured risk understanding: how faithfully participants represented the exact numbers displayed. Only assessing *comprehension* fails to capture individuals' *impressions* of plotted information (Stone et al., 2015). Understanding the 'gist' obtained from a visualisation is crucial since this takes precedence over 'verbatim' information when making decisions (Reyna, 2008). Reyna (2008) argues that assessing gist requires consideration of how the *absolute* magnitude of values is interpreted, since *relative* differences are only one aspect of a dataset conveyed by a visualisation.

Okan et al. (2018) collected magnitude ratings, yet these cannot be examined in isolation, since they were assimilated into a combined measure of perceived risk, along with ratings of the degree of *difference between* plotted values. Outside the risk communication literature, a substantial body of research has also demonstrated that axis range affects judgements of the difference between values (Correll et al., 2020; Pandey et al., 2015; Witt, 2019; Yang et al., 2021). In spite of this, research has typically neglected the effects of axis range on judgements of the magnitude of *the values themselves*. Sandman et al. (1994) observed that manipulating axis limits in risk ladders influenced absolute magnitude judgements, but did not use axes to convey denominators. This is the focus of the experiments presented in this work.

Prior work on extending axes did not disclose methods for determining axis limits in charts without denominators. The values used as upper limits appear to be arbitrary. In the present study, we increase ecological validity by employing default axis limits from *ggplot2* (Wickham, 2016), a popular visualisation tool used in the R programming environment. Furthermore, previous studies' statistical power and generalisability have been limited by the use of one (Okan et al., 2018) or two (Garcia-Retamero and Galesic, 2010) trials per participant. Our experiments explore a range of scenarios, with 32 experimental trials per participant. The data presented are also unrelated to risk judgements, the domain of prior research on this topic.

2.2.3 Open Research Statement

Pre-registrations, experiment scripts, materials, data, and analysis code are available at osf.io/. A Dockerfile is also included to facilitate reproduction of the computational environment used for analysis, allowing generation of a fully-reproducible version of this paper.

2.3 Experiment 1

2.3.1 Introduction

This experiment investigates the influence of axis limits on interpretations of plotted values' magnitude. Participants viewed bar charts with default axes, or axes which extended to a denominator value well above the bars. Comparing participants' interpretations captures the influence of displaying the same data with and without numerical context.

2.3.2 Method

2.3.2.1 Materials

We developed 40 scenarios about fictitious studies (e.g., a quality control study). Each fictitious study evaluated a specific outcome across five categories (e.g., the number of items produced without defects, for five manufacturing methods). The denominator (e.g., total number of items produced) was identical for each category.

We generated bar charts in R (R Core Team, 2022) using *ggplot2* (version 4.1.2), *tidyverse* (version 1.3.1) and *ggh4x* (version 0.2.1). The two versions of each chart displayed the same five values, but employed different y-axis limits. Denominator values (400, 500, or 600) were used to generate datasets: data were sampled from a normal distribution with a mean equal to either 20% or 40% of a given denominator value, and a standard deviation equal to 1% of the denominator value.

For charts with extended axes, the denominator value was used as the y-axis upper limit. The other charts used a y-axis upper limit which was dictated by *ggplot2*'s default axis settings. These settings automatically identify a set of convenient breaks for each dataset, then slightly extend the plot area, adding an additional 5% of the original axis range. In both conditions, a smaller expansion factor of 1% was applied to the lower axis limit, in order to eliminate visible space below the 0 baseline. Figure 2.1 shows example charts for both conditions.

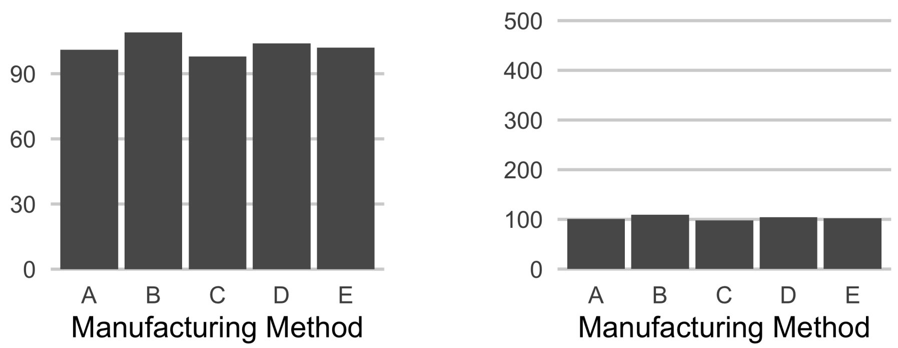

FIGURE 2.1: Example charts for Experiment 1. The same data appears in both charts. Accompanying text explained what the values represented: 'The graph shows, for each manufacturing method, how many of the items were free from defects'. The chart with the default axes (left) employs an upper limit determined by ggplot2. The chart with the extended axes (right) employs an upper limit equal to the dataset's denominator value.

For the majority of datasets generated, the default settings produced charts where the highest gridline did not exceed the tallest bar. For consistency, when the opposite situation occurred, we used a different random seed to generate an alternative dataset for both conditions. 10% of datasets used were generated using this method.

In experimental trials (32 total), plotted values consisted of relatively small proportions of the dataset's denominator value (roughly 20% or 40%). To introduce variety and encourage attention, eight filler trials showed plotted values which were roughly 90% of the corresponding denominator value. Denominators for filler trials were selected so that numerical labels on the y-axis would approximately resemble the y-axis labels of extended and default bar charts from experimental trials.

We included six attention check trials to assess participants' engagement with the task. These trials were similar to experimental and filler trials, consisting of text, a bar chart, a question and a visual analogue scale. However, participants were instructed to ignore the bar chart and provide a specified response on the visual analogue scale.

2.3.2.2 Design

We employed a within-participants design: participants viewed 16 different charts in each of the two conditions (32 experimental trials total). The correspondence between scenarios and conditions was counterbalanced using two lists. However, all participants saw the same versions of the eight filler items and six attention check items. There were a total of 46 trials, which were presented in a random order.

2.3.2.3 Procedure

We programmed the experiment using PsychoPy (version 2022.1.4, (Peirce et al., 2019)). Participants were instructed to carry out the experiment using a laptop or desktop computer (not a mobile phone or tablet). After providing informed consent, participants completed a demographic questionnaire and Garcia-Retamero et al.'s (2016) five-item subjective data visualisation literacy scale.

Participants were asked to imagine they were a researcher tasked with determining the outcome of experiments and surveys. They were instructed to provide an overall assessment of all data presented in a graph after studying the text, graph, and question. The questions asked about plotted values' magnitudes (e.g., 'How successful were the manufacturing methods?'), with participants responding on visual analogue scales with anchors at the extremes (e.g., 'Very unsuccessful', 'Very successful'). Figure 2.2 shows an example trial.

Participants were permitted to move the response marker as many times as they liked before proceeding to the next trial, but could not return to previous trials. The response scale's granularity was altered for each attention check item, such that participants could only respond at the extremes or the middle of the scale. Finally, participants were informed that all data presented was fictitious and were given the option to provide comments on

FIGURE 2.2: An example trial from Experiment 1, showing a bar chart with a default axis limit.

the experiment and describe any strategies used. Average completion time was 19.2 minutes.

2.3.2.4 Participants

Participants were recruited using Prolific.co. The experiment was advertised to fluent English speakers with normal-or-corrected to normal vision, who had previously participated in at least 100 studies on the site.

Data were returned by 157 participants. Per pre-registered exclusion criteria, seven participants' submissions were rejected because they answered more than one of six attention check questions incorrectly. Participants whose submissions were accepted received £3.50.

The final sample consisted of 150 participants (57.3% male, 39.3% female, 3.3% non-binary). Mean age was 33.18 years ($SD = 12.93$). The mean data visualisation literacy score was 21.35 ($SD = 4.73$), out of a maximum of 30.

This experiment was approved by the University of Manchester's Division of Neuroscience and Experimental Psychology Ethics Committee (ethics code: 2022-11115-24245).

2.3.3 Analysis

We conducted analysis using R (R Core Team, 2022) (version 4.2.1). Linear mixed models were built using *lme4* (Bates et al., 2015). Each initial model was based on a maximal model with random intercepts and slopes for participants and scenarios (Barr et al., 2013). The *buildmer* package (Voeten, 2022) was used to identify the final random effects structure, ensuring convergence and removing terms not significantly contributing to explaining variance.

2.3.3.1 Magnitude Ratings

Figure 2.3 shows the distribution of ratings for charts with default axes and extended axes.

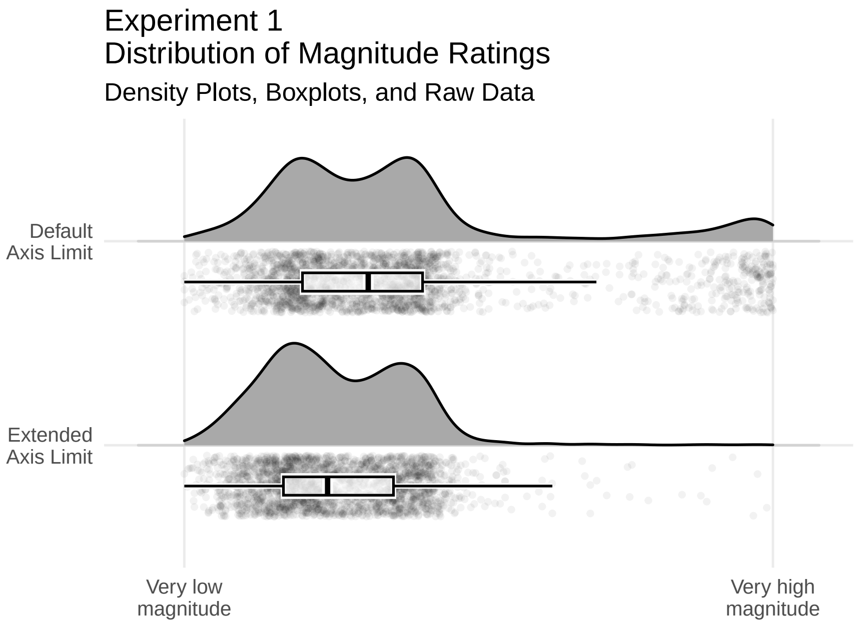

FIGURE 2.3: The distribution of visual analogue scale ratings in response to default and extended axis limits. Each circle represents a participant's response to an individual bar chart. Note charts with a default axis limit have a higher median value (represented by the line in the centre of the boxplot) and have a larger proportion of ratings near the 'Very high magnitude' end of the scale, compared to charts with an extended axis limit.

Linear mixed-effects modelling revealed that participants awarded higher ratings to charts with default axes, compared to charts with extended axes: $F(1, 152.54) = 44.90$, $p < .001$, partial $\eta^2 = 0.23$ (a large effect size).

This model employed a maximal random effects structure, capturing the baseline responses (intercepts) and differences between the two axis limits (slopes) separately for each individual participant and each individual scenario. The model formula was as follows: `magnitude ~ axis_limit + (1 + axis_limit | participant) + (1 + axis_limit | scenario)`.

2.3.3.2 Magnitude Ratings and Data Visualisation Literacy

Accounting for differences in data visualisation literacy did not change the significant effect of axis limits: $F(1, 152.57) = 44.95$, $p < .001$, partial $\eta^2 = 0.23$ (a large effect size). This model employed the same random effects structure as above. The model formula was as follows: `magnitude ~ axis_limit + literacy + (1 + axis_limit | participant) + (1 + axis_limit | scenario)`

However, an exploratory analysis (not specified in the pre-registration) revealed an interaction between data visualisation literacy and axis limits: $F(1, 146.75) = 11.74$, $p = .001$, partial $\eta^2 = 0.07$ (a medium effect size). The difference between magnitude ratings for data points presented with different axis limits diminished as data visualisation literacy increased. This is shown in Figure 2.4. This model employed random intercepts for participants, with random slopes for axis limits, plus random intercepts for items, with random slopes for axis limits. The model formula was as follows: `magnitude ~ axis_limit * literacy + (1 + axis_limit | participant) + (1 + axis_limit | scenario)`

2.3.4 Discussion

This experiment explored the consequences of including a graphical cue to a denominator value using bar charts' axes. In bar charts with extended axes, denominator information was available through the accompanying text *and* through the upper axis limit. Comparison with bar charts employing default axes, where denominator information was *only* available through text, reveals the contribution of a graphical cue to a denominator value. We observed that plotted values' magnitudes were interpreted as *greater* when a default axis limit was used, compared to an axis limit equal to the dataset's denominator value. Therefore, assessments of data were biased by the numerical context provided by axes.

In both conditions, there was a large proportion of magnitude ratings near the lower end of the rating scale. This indicates some attention to denominator information in the absence of a graphical cue. However, an increased presence of extreme high magnitude ratings in the condition with default axes suggests that the appearance of tall bars carried

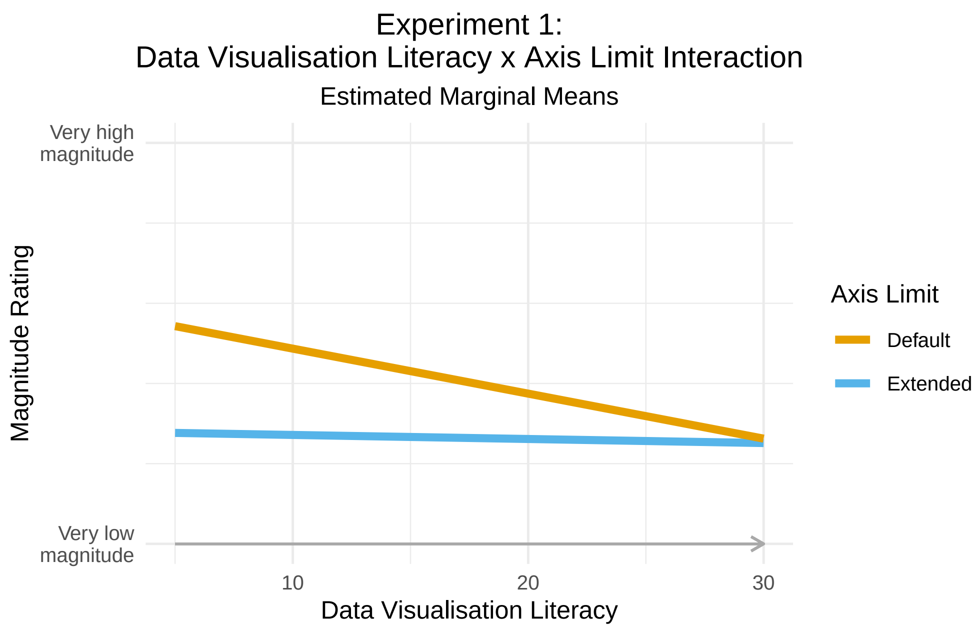

FIGURE 2.4: The interaction between data visualisation literacy and axis limits, for magnitude ratings in Experiment 1. The data displayed are based on the estimated marginal means from the inter- action model.

the implication of large values. These extreme ratings may indicate a failure to account for denominator information in the absence of a graphical cue. We investigate the role of denominator information further in Experiment 2.

2.4 Experiment 2

2.4.1 Introduction

Experiment 1 observed differences in interpretations of data presented using different axes limits. Overall, plotted data were associated with lower magnitudes when presenting using axes which extended to a denominator value. Compared to bar charts with extended axes, charts with no graphical cue to a denominator value elicited a wider range of responses. This variety appears to reflect differences in how the denominator information supplied in accompanying text is used in magnitude judgements. This raises questions about how external sources of denominator information might influence the interpretation of different chart designs.

In Experiment 2, we manipulate the presence of denominator information in accompanying text, in addition to manipulating axis limits. This experiment investigates how these textual and graphical cues inform assessments of data. This can reveal how different chart designs are interpreted with and without additional cues to numerical context. This 2x2 design provides an opportunity to replicate the primary finding from Experiment 1 and can also examine whether responses in the absence of denominator information resembles the previously observed pattern of extreme ratings.

This second experiment requires minor adaptations to materials and procedure. First, highly ambiguous trials without denominator information supplied in text may elicit unreliable random responses. Therefore, we collect additional *confidence ratings* to directly index this aspect of participants' evaluations. This provides a more comprehensive view of participants' cognitive representations. Second, when denominators are not supplied in text, participants may use denominator values supplied in previous trials to inform their judgements. A limited range of denominators (as in Experiment 1) would artificially diminish uncertainty regarding possible values, inhibiting authentic, spontaneous judgements. Therefore, we expand the range of denominator values in Experiment 2. Third, increasing the number of fillers (which depict relatively high magnitudes) to match the number of experimental items (which depict relatively low magnitudes) will avoid priming effects by ensuring high and low magnitudes seem equally plausible.

2.4.2 Method

2.4.2.1 Materials

We generated bar charts in R using *ggplot2* (version 4.2.1), *tidyverse* (version 1.3.2) and *ggh4x* (version 0.2.3).

Bar charts were generated using the same method as in Experiment 1. We used the same scenarios from Experiment 1, and generated 24 new scenarios for use as additional filler items, thus employing 32 experimental items and 32 filler items. To increase variation across datasets, we employed a wider range of denominators (200, 400, 600, and 800) meaning the plotted values for each scenario also differed from Experiment 1.

We added the word 'surveyed' or 'assessed' to the accompanying text for seven items where the absence of a denominator may have implied that data were collected for the entire population under study. For example, where the fictitious study concerned data collected in five towns, the final sentence read 'The graph shows, for each town, how many people *surveyed* used public transport regularly', to avoid the implication that the denominator was equal to an entire town's population. This ensured that the inclusion of denominator values was equally informative across all scenarios.

16% of datasets used were re-generated to ensure that the highest gridline of a default axis did not exceed the highest plotted value.

2.4.2.2 Design

We employed a within-participants 2x2 Latin-squared design with two factors: axis limits (default vs. extended) and denominator information (present vs. absent). Participants viewed 8 different charts for each combination of conditions (32 experimental trials total). The correspondence between scenarios and conditions was counterbalanced using four lists. However, all participants saw the same versions of the 32 filler items and six attention check items.

2.4.2.3 Participants

Participants were recruited using Prolific.co, using the same inclusion criteria as Experiment 1. Additionally, the experiment was not advertised to individuals who completed Experiment 1.

Data were returned by 208 participants. Per pre-registered exclusion criteria, eight participants' submissions were rejected because they answered more than one of six attention check questions incorrectly. Participants whose submissions were accepted received £5.00.

The final sample consisted of 200 participants (60% male, 38% female, 1% non-binary, 0.5% other, 0.5% prefer not to say). Mean age was 33.17 years (SD = 10.34)[1]. The mean data visualisation literacy score was 21.72 (SD = 4.84), out of a maximum of 30.

This experiment was approved by the University of Manchester's Division of Neuroscience and Experimental Psychology Ethics Committee (ethics code: 2023-11115-28428).

2.4.2.4 Procedure

The procedure was identical to Experiment 1, except for the addition of a confidence rating, where participants were asked 'How confident are you in your response?'. The anchors on the response scale were 'Not very confident' and 'Very confident'. Figure 2.5 shows an example trial.

[1]Age data were unavailable for 2 participants

For attention check items, participants were asked to provide a specific response on the magnitude rating scale, and a random response on the confidence rating scale.

Average completion time was 29.3 minutes.

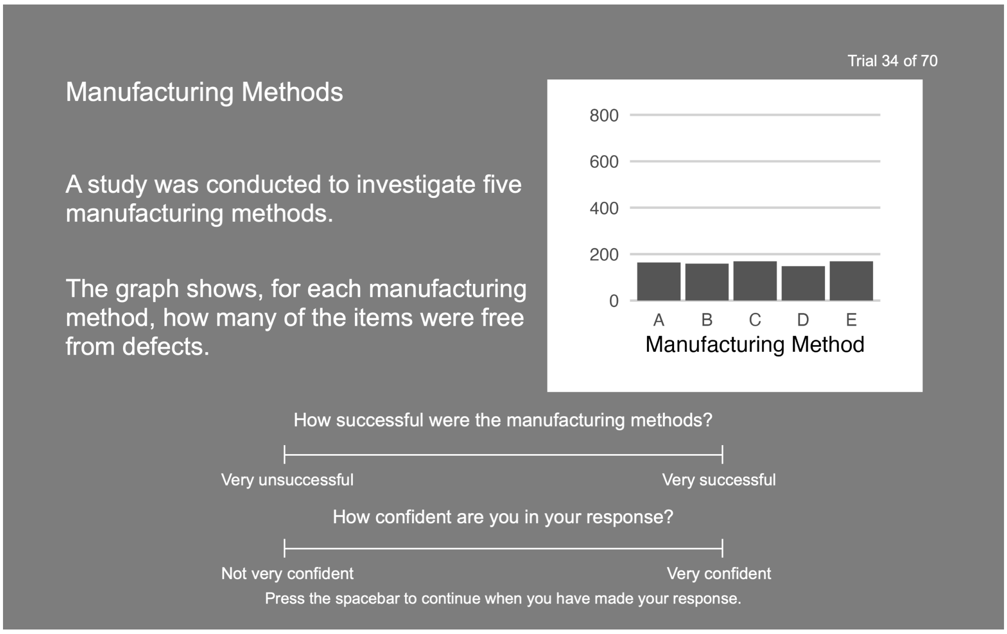

FIGURE 2.5: An example trial from Experiment 2, showing a bar chart with an extended axis limit. Note the presence of an additional confidence rating scale.

2.4.3 Analysis

2.4.3.1 Magnitude Ratings

Figure 2.6 shows the distribution of magnitude ratings for charts with default axes and extended axes, where denominators were absent from text, and where they were present.

A mixed effects model revealed that charts with default axes elicited higher ratings compared to charts with extended axes (F(1, 198.22) = 311.45, $p < .001$, partial $\eta^2 = 0.61$, a large effect size) and charts not accompanied by a denominator in text elicited higher ratings than those accompanied by a denominator (F(1, 82.23) = 380.50, $p < .001$, partial $\eta^2 = 0.82$, a large effect size).

Crucially, there was also a significant interaction between axis limits and denominator information: F(1, 5,741.16) = 1,540.86, $p < .001$, partial $\eta^2 = 0.21$ (a large effect size). Figure 2.7 plots this interaction.

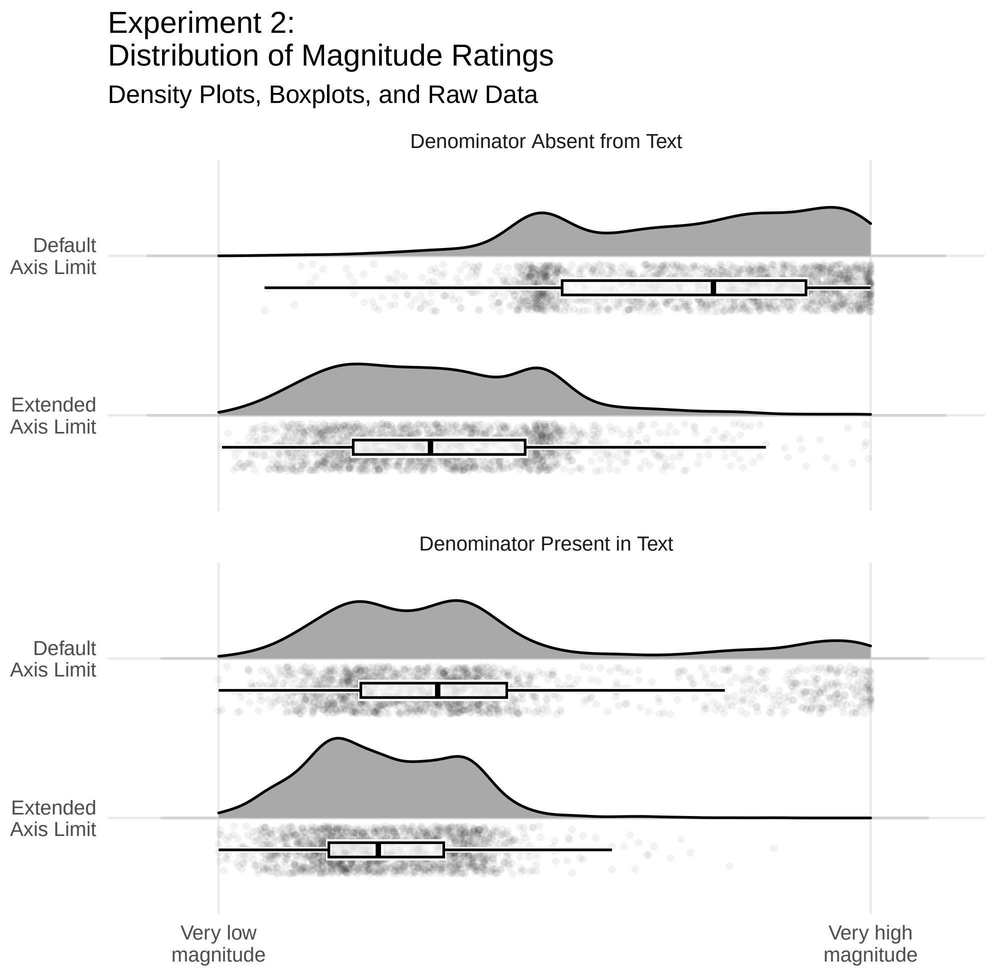

FIGURE 2.6: The distribution of magnitude ratings in response to default and extended axis limits, shown separately for trials where denominator values were *absent from* accompanying text (top), and trials where denominator values were *present in* accompanying text (bottom). Each circle represents a participant's response to an individual bar chart.

Pairwise comparisons conducted using *emmeans* (Lenth, 2021) revealed that charts with extended and default axes were rated differently when the denominator was present, replicating the effect from Experiment 1 (z = -9.19, $p < .001$, Cohen's *d* = 0.33, a small effect size), and also when the denominator was absent (z = -25.35, $p < .001$, Cohen's *d* = 0.90, a large effect size). Therefore, the interaction indicates that the *degree* to which a bar chart's axis affected interpretations varied according to whether the denominator was present or absent.

This model employed random intercepts for each participant, with random slopes for axis limits and denominator information, plus random intercepts for each scenario, with random slopes for denominator information. The model formula was as follows: `magnitude ~ axis_limit * denominator + (1 + axis_limit + denominator | participant) + (1 + denominator | scenario)`

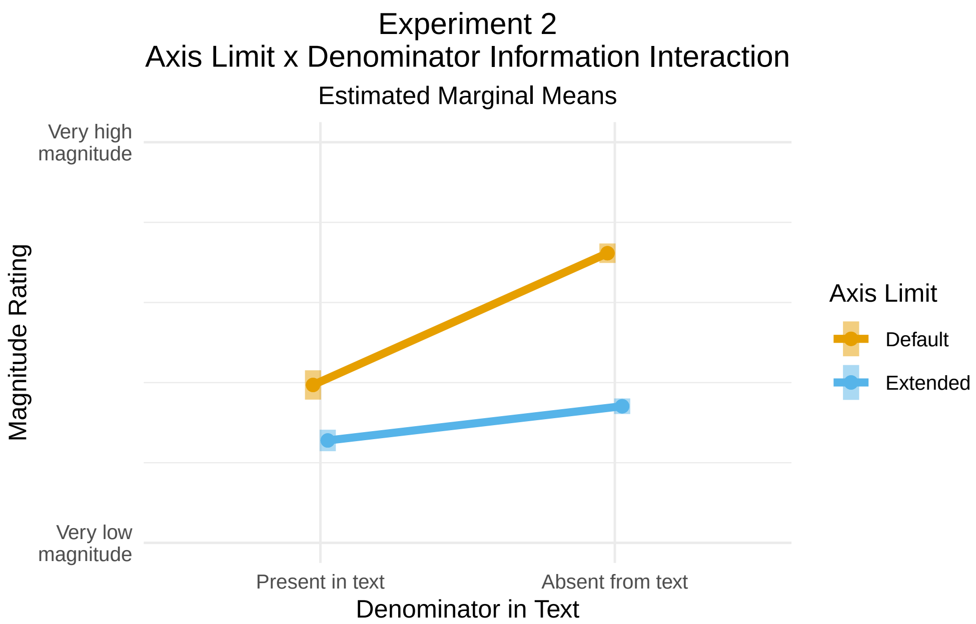

FIGURE 2.7: The interaction between axis limits and denominator information, for magnitude ratings. Estimated marginal means are generated by the linear mixed model used in analysis. Translucent bars show 95% confidence intervals.

2.4.3.2 Magnitude Ratings and Data Visualisation Literacy

Accounting for differences in data visualisation literacy did not eliminate the main effect of axis limits ($F(1, 198.24) = 311.46$, $p < .001$, partial $\eta^2 = 0.61$, a large effect size), the main effect of denominator information ($F(1, 82.29) = 380.60$, $p < .001$, partial $\eta^2 = 0.82$, a large effect size), or the significant interaction: $F(1, 5{,}741.14) = 1{,}540.85$,

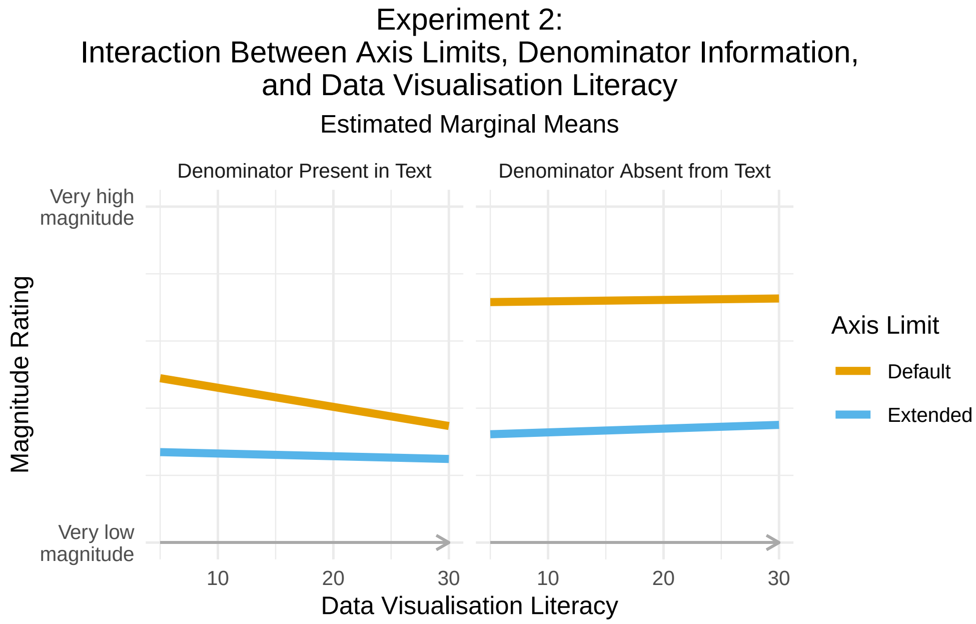

FIGURE 2.8: Three-way interaction between axis limits, denominator information, and data visualisation literacy, for magnitude ratings in Experiment 2. The data displayed are based on the estimated marginal means from the interaction model. Note that this interaction is associated with a very small effect size, driven by the relationship between magnitude ratings and data visualisation literacy for bar charts with default axes where denominators were present in text.

p $<$ .001, partial η^2 = 0.21, a large effect size). This model employed the same random effects structure as above. The model formula was as follows: `magnitude ~ literacy + axis_limit * denominator + (1 + axis_limit + denominator | participant) + (1 + denominator | scenario)`

However, an exploratory analysis revealed a significant three-way interaction between data visualisation literacy, axis limits, and denominator information: F(1, 5,768.98) = 10.39, p = .001, partial $\eta^2 < .01$ (a very small effect size). Pairwise comparison reveals that differences in data visualisation literacy only predicted magnitude ratings for charts with default axes where the denominator was present in text. For this combination of conditions, magnitude ratings were lower for participants with higher data visualisation scores: z = -1.98, p = .048, Cohen's *d* = 0.07, (a very small effect size). There were no other significant changes in magnitude ratings as a function of data visualisation literacy, for charts with extended axes where the denominator was present in text (z = -0.69, p = .490, Cohen's *d* = 0.02, a very small effect size), for charts with default axes where the denominator was absent from text (z = 0.19, p = .846, Cohen's *d* = 0.01, a very small effect size), or for charts with extended axes where the denominator was absent from text (z = 0.72, p = .472, Cohen's $d = 0.03$, a very small effect size). This three-way interaction is shown in Figure 2.8. This model employed random intercepts for participants, with random slopes for axis limits and denominator information, plus random slopes for scenarios. The model formula was as follows: `magnitude ~ axis_limit * denominator * literacy + (1 + axis_limit + denominator | participant) + (1 | scenario)`.

2.4.3.3 Confidence Ratings

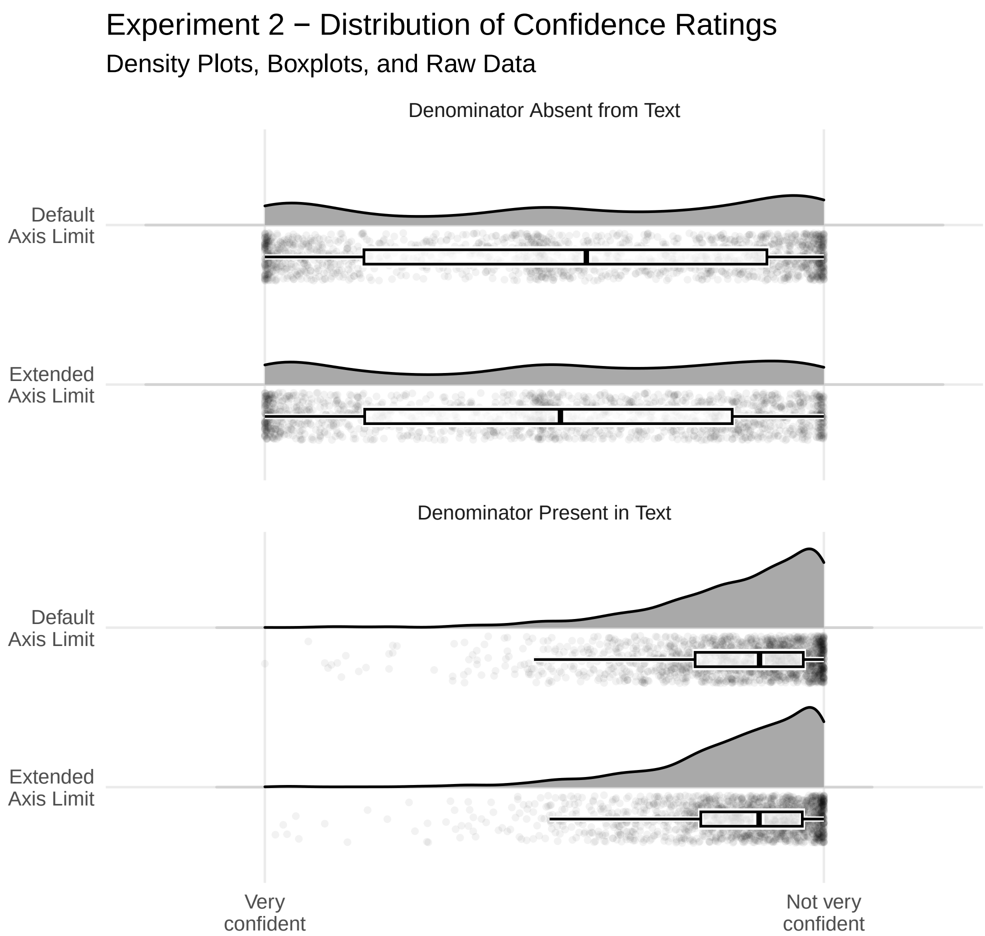

FIGURE 2.9: The distribution of confidence ratings in response to default and extended axis limits, shown separately for trials where denominator values were *absent from* accompanying text (top), and trials where denominator values were *present in* accompanying text (bottom). Each circle represents a participant's response to an individual bar chart.

Figure 2.9 shows the distribution of confidence ratings for charts with default axes and extended axes, where denominators were absent from text, and where they were present.

A mixed effects model revealed a main effect associated with axis limits (F(1, 199.00) = 5.97, p = .015, partial $\eta^2 = 0.03$, a small effect size); a main effect associated with

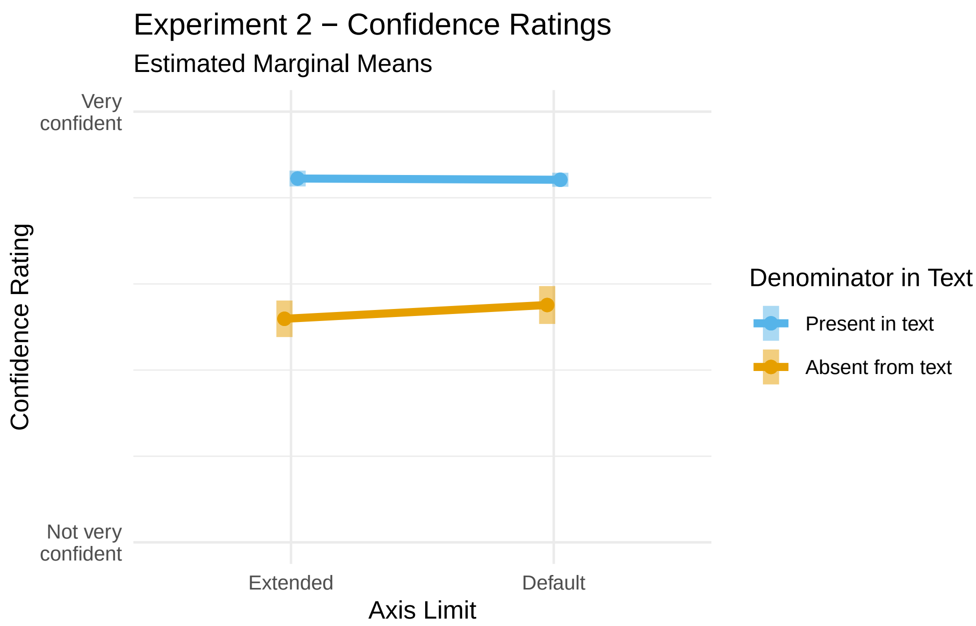

FIGURE 2.10: The interaction between axis limits and denominator information, for confidence ratings. Estimated marginal means are generated by the linear mixed model used in analysis. Translucent bars show 95% confidence intervals.

denominator information (F(1, 198.99) = 184.93, p < .001, partial η^2 = 0.48, a large effect size); and an interaction between axis limits and denominator information (F(1, 5,799.00) = 27.74, p < .001, partial η^2 < .01, a very small effect size). This interaction consisted of a difference between extended and default charts when the denominator was absent from text (z = -4.69, p < .001, Cohen's *d* = 0.17, a very small effect size), but no difference between charts when the denominator was present (z = 0.42, p = .988, Cohen's *d* = 0.01, a very small effect size). However, it is clear from Figure 2.10, as well as the partial η^2 values, that the effect sizes associated with both axis limits and the interaction are trivial.

This model employed random intercepts for participants with random slopes for axis limits and denominator information. The model formula was as follows: `confidence ~ denominator * axis_limit + (1 + denominator + axis_limit | participant)`

2.4.3.4 Confidence Ratings and Data Visualisation Literacy

Accounting for differences in data visualisation literacy did not change the pattern of results for confidence ratings. There was a main effect associated with axis limits (F(1, 199.01) = 5.97, $p = .015$, partial $\eta^2 = 0.03$, a small effect size); a main effect associated with denominator information ($F(1, 198.99) = 184.93$, $p < .001$, partial $\eta^2 = 0.48$, a large effect size); and an interaction between axis limits and denominator information ($F(1, 5{,}798.99) = 27.74$, $p < .001$, partial $\eta^2 < .01$, a very small effect size). This model employed the same random effects structure as above. The model formula was as follows: `confidence ~ literacy + denominator * axis_limit + (1 + denominator + axis_limit | participant)`

An exploratory analysis revealed that there was no interaction between data visualisation literacy and axis limits for confidence ratings ($F(1, 198.00) = 1.04$, $p = .310$, partial $\eta^2 < .01$, a very small effect size); no interaction between data visualisation literacy and denominator information ($F(1, 197.99) = 2.95$, $p = .087$, partial $\eta^2 = 0.01$, a small effect size); and no three-way interaction between data visualisation literacy, axis limits, and denominator information ($F(1, 197.99) = 0.16$, $p = .685$, partial $\eta^2 < .01$, a very small effect size). This model employed random intercepts for participants, with random slopes for axis limits, denominator information, and the interaction between axis limits and denominator information. The model formula was as follows: `confidence ~ axis_limit * denominator * literacy + (1 + axis_limit * denominator | participant)`

2.4.4 Discussion

This experiment manipulated bar charts' axis limits and the presence of denominator values in accompanying text. The results demonstrate that values presented in charts with default axis limits are associated with higher magnitude judgements than charts with extended axes, in both the presence and absence of denominator information. However, the absence of denominator information amplifies this bias. These results also suggest that extreme high magnitude ratings for default charts in the presence of a denominator value may be driven by a failure to incorporate that value into reasoning. Finally, confidence in judgements is reliably affected by the inclusion of denominator information in text.

2.5 General Discussion

Axis limits can be easily manipulated in common data visualisation software, in order to include a visual cue to denominator information. However, defaults axis limits are often based on plotted data only, so typically omit denominator information. We demonstrate that the magnitude of plotted values was interpreted as smaller in bar charts with axes that extended to the denominator value, compared to those employing default axis settings. The influence of axis limits was particularly large when no denominator information was included in the text accompanying a chart. This provides insight into the cognitive process involved in magnitude judgements, indicating that denominator information is an important aspect in interpreting absolute magnitude in bar charts.

In Experiment 1, we identified a framing effect, wherein charts with axes that accommodated a denominator value elicited smaller magnitude judgements compared to charts with default axes. In both conditions, the denominator was explicitly presented in the text. However, some extreme responses in the default condition appeared disregard the denominator information. We conducted another experiment in order to examine the role of the denominator in the cognitive processing of magnitude. In Experiment 2, we examined how interpretations were affected by the absence of denominator information, thus capturing how this information was incorporated differently across chart designs.

Experiment 2 makes several additional contributions. First, this experiment replicated the main effect from Experiment 1. That is, we observed that magnitudes were interpreted as smaller when values were shown with an extended axis, rather than a default axis. Second, the experiment illustrated the impact of including the denominator in accompanying text. This cue affects viewers' interpretations differently depending on whether a chart's axis also incorporates the same value. Without denominator information in text, the magnitude of values plotted using default axes can be ambiguous. Accordingly, drastically higher ratings in the absence of denominator information illustrate the denominator's role in reducing ambiguity. Interpretation of values plotted using extended axes was affected to a lesser extent by the denominator's absence. Thus, the impact of a bar chart's axis is smaller when it is accompanied by a denominator. This suggests axis limits provide a visual cue to denominator information when interpreting magnitudes.

Third, Experiment 2 replicated the pattern of responses observed in Experiment 1 for charts with default axes and accompanying denominator information. This pattern consists of a small number of higher magnitude ratings, in contrast to the general tendency for lower magnitude ratings. Figure 2.6 reveals a close resemblance between the distribution of these higher ratings and the overall distribution of ratings for default charts *without* accompanying denominator information. This suggests that these extreme ratings may share a cause. Unusually high responses in the presence of denominator information likely result from failure to account for the denominator and a subsequent reliance on the chart's appearance. The analogous responses to charts without accompanying denominators (Experiment 2) can be considered an experimentally-induced instance of the same effect.

Fourth, additional ratings collected in Experiment 2 provide insight into participants' confidence. Although analysis of these ratings indicated a main effect of axis limits and an interaction between denominator information and axis limits, the minuscule effect sizes cast doubt over the practical significance of these effects. In spite of this, absence of a denominator clearly lowered confidence. This suggests that participants were hesitant to form magnitude judgements based solely on a bar chart's appearance. Inclusion of a denominator value in text was preferred regardless of graphical cues to context.

2.5.1 Relationship to Prior Work

Our focus on judgements of values' magnitudes is noteworthy because the vast majority of related work has explored participants' judgements of *differences between* values (Correll et al., 2020; Driessen et al., 2022; Garcia-Retamero and Galesic, 2010; Okan et al., 2020, 2018, 2012; Pandey et al., 2015; Stone et al., 2018, 2003; Witt, 2019; Yang et al., 2021). Responses to questions about values' magnitudes have often been obscured through inclusion in composite measures (e.g., Okan et al., 2018), or have been collected to assess comprehension, rather than interpretation (e.g., Garcia-Retamero and Galesic, 2010). As Stone et al. (2015) discuss, failing to consider interpretations of values' magnitudes reflects two issues. First, neglecting values' magnitudes overlooks a relevant aspect of numerical information. Second, neglecting participants' *interpretations* limits insight into decision-making, which is not simply governed by accurate retrieval of information (see Reyna, 2008).

Whereas much prior research has been limited to interpretation of risk information (Garcia-Retamero and Galesic, 2010; Okan et al., 2020, 2018, 2012; Stone et al., 2018, 2017, 2015, 2003), we demonstrate that biases in interpretation extend to a wide range of non-risk scenarios. This provides confidence that these findings are widely applicable, and using *multiple* trials per participant enhances statistical power. When generating charts that do not include denominator values, previous experiments (Garcia-Retamero and Galesic, 2010; Okan et al., 2018) appear to have employed arbitrary axis limits. By employing axis limits based on *ggplot2*'s default settings, our materials reflect a common practice, enhancing our experiment's ecological validity.

We contribute to a large body of evidence illustrating biases in the interpretation of numerical information, specifically *framing effects* (Tversky and Kahneman, 1981). Our results are consistent with research demonstrating that manipulating bar charts' axis limits influences interpretation of plotted values (Garcia-Retamero and Galesic, 2010; Okan et al., 2018). Furthermore, Okan et al. (2018) found that participants' perceptions of risk were influenced more by bar charts' axis limits when labels containing numerators and denominators were excluded. Similarly, we observed that interpretations of magnitude were influenced more by bar charts' axis limits when denominator values were omitted from accompanying text.

A previous study exploring interpretation of magnitude in bar charts observed different responses according to whether stacked bars or blank space conveyed alternative outcomes (Stone et al., 2017). We demonstrate that manipulating the amount of blank space above bars can elicit different magnitude judgements, without plotting alternative outcomes explicitly. Earlier work investigating (in)consistency in the formats used to display numerators and denominators is also relevant. Stone et al. (2015) found that displaying a value using icons, accompanied by a denominator in text, increased impressions of that value's magnitude, compared to when both the value and denominator were presented in text. We too found higher ratings when values displayed using bars were only accompanied by a denominator in text, compared to when a corresponding graphical cue to the denominator value was also present.

According to Fuzzy Trace Theory, different interpretations can arise due to different gist-level representations, despite accurate comprehension of presented values (Reyna, 2008). Therefore, access to denominator information in accompanying text did not prevent our chart designs influencing judgements. Encoding of gist is reported to be influenced by the appearance of graphical elements (Reyna, 2008). This suggests that the taller bars in our default axis conditions were responsible for impressions of greater magnitude, compared to shorter bars in our extended axis conditions. That charts with extended axes elicited lower magnitude ratings is also consistent with Stone et al.'s (2018) proportional reasoning account, which suggests that part-to-whole displays facilitate processing of a larger numerical context.

We did not find evidence that data visualisation literacy affected our results. This is contrary to the finding that data visualisation literacy predicted the efficacy of using icon arrays to reduce denominator neglect (Okan et al., 2012). However, this is consistent with the finding that the impact of manipulations like ours (axis range, numerical labels) are independent of data visualisation literacy (Okan et al., 2018). This measure may capture whether people have sufficient ability to extract information from a visualisation, rather than predicting the degree to which they will be influenced by subtler design choices (Yang et al., 2021). Numeracy is associated with decreased sensitivity to framing effects (Peters et al., 2006), so this may be a better candidate for understanding individual differences in response to visualisation design.

2.5.2 Implications

When conveying values' magnitudes, both axis limits and accompanying text warrant consideration from data visualisation designers. A bar chart produced using default settings is not equivalent to a bar chart with an axis that incorporates a denominator value. Extending an axis in this manner increases consistency in judgements and may provide insurance against individuals who fail to account for accompanying denominator information. Similarly, where constraints prevent inclusion of a denominator value in text, an extended axis should facilitate viewers' recognition of this numerical context. We observed that confidence ratings were consistently high in the absence of a denominator in text, despite use of an extended axis. Explicitly providing denominator values in text, regardless of graphical cues, would therefore promote viewers' confidence in their judgements.

It is also worth considering situations which may accentuate the observed bias. High cognitive load exacerbates the numerosity bias (Pelham et al., 1994), therefore may also interfere with magnitude judgements. Even when denominator information is supplied in text, high cognitive load could prevent this information from informing interpretations. This would likely increase reliance on bar charts' appearances, like in Experiment 2. Additionally, assuming that an audience has knowledge of a dataset's denominator may increase biases in individuals who are unfamiliar with the topic.

2.5.3 Limitations and Future Work

This work is concerned with visualisations intended to convey plotted values' magnitudes. However, design considerations will differ when conveying *differences* between values. In this case, axis ranges should be determined by the magnitude of the differences (Correll et al., 2020; Witt, 2019; Yang et al., 2021). Consequently, our recommendations are not relevant for all communicative scenarios. However, maintaining awareness of the implication of plotted values' magnitudes may help avoid misinterpretation of data, even if this type of judgement is not a primary concern.

Our experiments apply best to controlled scenarios, such as surveys and experiments where all plotted values share the same denominator. These findings may also extend to datasets with unequal denominators, if bars are used to depict proportions or percentages, permitting use of a single meaningful axis limit. However, this design will not be suitable for plotting other types of dataset. We also acknowledge that proportions are not the only factor influencing magnitude judgements: subject matter is also likely to inform assessments. For example, bars clearly depicting one or two hours spent on administrative tasks within a 35-hour work week will still elicit some differences of opinion regarding whether these values are high or low.

All materials were produced using *ggplot2*. Therefore, our conclusions about default axis limits only pertain to bar charts created using this package's settings, though we expect other visualisation libraries' default settings to elicit similar responses, due to similarity in their behaviour. For uniformity in our materials, we only employed default charts where the highest gridline was positioned below the highest value, since this was the most common visual arrangement. We did not examine the minority of cases where the highest gridline exceeds the highest value. Whether this influences magnitude judgements could be explored in future experiments. In addition, future work should employ decision-making tasks to quantify the impact of axis limits on applied judgements.

2.5.4 Conclusion

In two experiments, we generated evidence on the effects of default and extended axis limits, illustrating the influential role of denominators in gauging magnitude. We provide insight into the cognitive processes involved in interpreting plotted values' magnitudes in bar charts and offer recommendations for facilitating judgements. Framing effects demonstrate the power of presentation choices on the interpretation of numbers.

References

Barr DJ, Levy R, Scheepers C, Tily HJ. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**:255–278. doi:10.1016/j.jml.2012.11.001

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software* **67**. doi:10.18637/jss.v067.i01

Burson KA, Larrick RP, Lynch JG. 2009. Six of One, Half Dozen of the Other: Expanding and Contracting Numerical Dimensions Produces Preference Reversals. *Psychological Science* **20**:1074–1078. doi:10.1111/j.1467-9280.2009.02394.x

Correll M, Bertini E, Franconeri S. 2020. Truncating the Y-Axis: Threat or Menace?Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Honolulu HI USA: ACM. pp. 1–12. doi:10.1145/3313831.3376222

Cosmides L, Tooby J. 1996. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* **58**:1–73. doi:10.1016/0010-0277(95)00664-8

Driessen JEP, Vos DAC, Smeets I, Albers CJ. 2022. Misleading graphs in context: Less misleading than expected. *PLOS ONE* **17**:e0265823. doi:10.1371/journal.pone.0265823

Garcia-Retamero R, Cokely ET, Ghazal S, Joeris A. 2016. Measuring Graph Literacy without a Test: A Brief Subjective Assessment. *Medical Decision Making* **36**:854–867. doi:10.1177/0272989X16655334

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

Lenth RV. 2021. Emmeans: Estimated Marginal Means, aka Least-Squares Means.

Levin IP. 1987. Associative effects of information framing. *Bulletin of the Psychonomic Society* **25**:85–86. doi:10.3758/BF03330291

Li M, Chapman GB. 2013. A big fish or a small pond? Framing effects in percentages. *Organizational Behavior and Human Decision Processes* **122**:190–199. doi:10.1016/j.obhdp.2013.07.003

Monga A, Bagchi R. 2012. Years, Months, and Days versus 1, 12, and 365: The Influence of Units versus Numbers. *Journal of Consumer Research* **39**:185–198. doi:10.1086/662039

Okan Y, Garcia-Retamero R, Cokely ET, Maldonado A. 2012. Individual Differences in

Graph Literacy: Overcoming Denominator Neglect in Risk Comprehension. *Journal of Behavioral Decision Making* **25**:390–401. doi:10.1002/bdm.751

Okan Y, Stone ER, Bruine De Bruin W. 2018. Designing Graphs that Promote Both Risk Understanding and Behavior Change. *Risk Analysis* **38**:929–946. doi:10.1111/risa.12895

Okan Y, Stone ER, Parillo J, Bruine de Bruin W, Parker AM. 2020. Probability Size Matters: The Effect of Foreground-Only versus Foreground+Background Graphs on Risk Aversion Diminishes with Larger Probabilities. *Risk Analysis* **40**:771–788. doi:10.1111/risa.13431

Pandey AV, Rall K, Satterthwaite ML, Nov O, Bertini E. 2015. How Deceptive are Deceptive Visualizations?: An Empirical Analysis of Common Distortion TechniquesProceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15. Seoul, Republic of Korea: ACM Press. pp. 1469–1478. doi:10.1145/2702123.2702608

Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, Kastman E, Lindeløv JK. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* **51**:195–203. doi:10.3758/s13428-018-01193-y

Pelham BW, Sumarta TT, Myaskovsky L. 1994. The Easy Path From Many To Much: The Numerosity Heuristic. *Cognitive Psychology* **26**:103–133. doi:10.1006/cogp.1994.1004

Peters E, Västfjäll D, Slovic P, Mertz CK, Mazzocco K, Dickert S. 2006. Numeracy and Decision Making **17**.

R Core Team. 2022. R: A Language and Environment for Statistical Computing.

Reyna VF. 2008. A Theory of Medical Decision Making and Health: Fuzzy Trace Theory. *Medical Decision Making* **28**:850–865. doi:10.1177/0272989X08327066

Reyna VF, Brainerd CJ. 2008. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences* **18**:89–107. doi:10.1016/j.lindif.2007.03.011

Sandman PM, Weinstein ND, Miller P. 1994. High Risk or Low: How Location on a "Risk Ladder" Affects Perceived Risk. *Risk Analysis* **14**:35–45. doi:10.1111/j.1539-6924.1994.tb00026.x

Sevilla J, Isaac MS, Bagchi R. 2018. Format Neglect: How the Use of Numerical Versus Percentage Rank Claims Influences Consumer Judgments. *Journal of Marketing*

82:150–164. doi:10.1177/0022242918805455

Stone ER, Bruin W, Wilkins AM, Boker EM, MacDonald Gibson J. 2017. Designing Graphs to Communicate Risks: Understanding How the Choice of Graphical Format Influences Decision Making. *Risk Analysis* **37**:612–628. doi:10.1111/risa.12660

Stone ER, Gabard AR, Groves AE, Lipkus IM. 2015. Effects of Numerical Versus Foreground-Only Icon Displays on Understanding of Risk Magnitudes. *Journal of Health Communication* **20**:1230–1241. doi:10.1080/10810730.2015.1018594

Stone ER, Reeder EC, Parillo J, Long C, Walb L. 2018. Salience Versus Proportional Reasoning: Rethinking the Mechanism Behind Graphical Display Effects: Proportional Reasoning. *Journal of Behavioral Decision Making* **31**:473–486. doi:10.1002/bdm.2051

Stone ER, Sieck WR, Bull BE, Frank Yates J, Parks SC, Rush CJ. 2003. Foreground:background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational Behavior and Human Decision Processes* **90**:19–36. doi:10.1016/S0749-5978(03)00003-7

Tversky A, Kahneman D. 1981. The Framing of Decisions and the Psychology of Choice. *Science* **211**:453–458. doi:10.1126/science.7455683

Voeten CC. 2022. Buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects.

Wickham H. 2016. ggplot2. New York, NY: Springer Science+Business Media, LLC.

Witt JK. 2019. Graph Construction: An Empirical Investigation on Setting the Range of the Y-Axis. *Meta-Psychology* **2**:1–20. doi:10.15626/MP.2018.895

Yang BW, Vargas Restrepo C, Stanley ML, Marsh EJ. 2021. Truncating Bar Graphs Persistently Misleads Viewers. *Journal of Applied Research in Memory and Cognition* S2211368120300978. doi:10.1016/j.jarmac.2020.10.002

Zhang YC, Schwarz N. 2012. How and Why 1 Year Differs from 365 Days: A Conversational Logic Analysis of Inferences from the Granularity of Quantitative Expressions. *Journal of Consumer Research* **39**:248–259. doi:10.1086/662612