Magnitude Judgements Are Influenced by Data Points' Relative Positions Within Axis Limits

Abstract

When visualising data, chart designers have the freedom to choose the upper and lower limits of charts' numerical axes. Axis limits can determine the physical characteristics of plotted values, such as the physical position of data points in dot plots. In three experiments (total N=420), we demonstrate that axis limits affect viewers' interpretations of the magnitudes of plotted values. Participants did not simply associate values presented at higher vertical positions with greater magnitudes. Instead, participants considered data points' relative numerical positions within the axis limits. Data points were considered to represent larger values when they were closer to the end of the axis associated with greater values, even when they were presented at the *bottom* of a chart. This provides further evidence of framing effects in the display of data, and offers insight into the cognitive mechanisms involved in assessing magnitude in data visualisations.

Introduction

Context is crucial for effectively judging the magnitude of numbers. A 10% probability is twice as great as a 5% probability, but in the absence of context, it is unclear whether this value should be considered large or small. When referring to the chance of losing one's job, a 10% probability may be considered large, but when referring to the chance of losing a sports bet, a 10% probability may be considered small.

Contextual cues may influence interpretation of magnitude in data visualisations. One such cue is the range of values on an axis, which can serve as a frame of reference for assessing whether a data point represents a large or small number. Figure 1 (a reproduction of a similar bar chart from the New York Times), which plots over time the number of Black members of the U.S. senate, provides a striking illustration. Unusually, the y-axis does not terminate just above the highest plotted value. Instead, the y-axis extends all the way to the maximum possible number of senators: 100. As a result, bars representing Black senators are confined to the very bottom, visible just above the x-axis, and a significant expanse of blank space looms above them. This framing situates plotted data points in their numerical context, thus conveying small magnitude.

It is unclear exactly how a viewer's inferences about magnitude might be influenced by axis range. Different axis limits present data points at different positions, so one possible explanation is that viewers interpret the magnitude of data points at higher positions as 'high' and those at lower positions as 'low'. Alternatively, axis limits may provide numerical context: plotted values may be judged as small in magnitude when the potential for larger values is clearly displayed. The present pair of experiments demonstrates the influence of axis limits on viewers' interpretations and explores which of these two accounts best explains how axis limits contribute to the communication of magnitude.

Overview

In three experiments, we manipulated the axis limits surrounding plotted data. The same data points either appeared close to the upper end of an axis range, or close to the lower end. Likert scale ratings of values' magnitudes were higher when data points were positioned close to the end of the axis which was associated with higher numbers. By employing charts with conventional and inverted y-axis orientations to distinguish between possible explanations, we reveal that magnitude judgements are influenced by data points' relative positions within the axis limits.

Related Work

Effects of Axis Limits on Comparison Judgements

Several studies have explored the role of axis limits in data visualisation. Research has typically focused on how axis limits can alter impressions of the *difference between* presented values. For example, when axis ranges are expanded to create blank space around a cluster of data points, correlation between those points is judged as stronger (Cleveland, Diaconis, and McGill 1982). Participants also rate the differences between values in bar charts as greater when the vertical gap between bars is larger due to a truncated y-axis (Pandey et al. 2015).

Correll et al.'s (Correll, Bertini, and Franconeri 2020) experiments found that greater truncation resulted in higher effect-size judgements in both line charts and bar charts. They found no reduction in effect size judgements when truncation was communicated using graphical techniques (e.g., axis breaks and gradients). Truncation effects also persisted even when participants estimated the values of specific data points. This suggests the bias is driven by initial impressions, rather than by a misinterpretation of the values portrayed by graphical markings.

The unavoidable consequence, Correll et al. suggest, is that designers' choices will influence viewers' interpretations whether axes are truncated or not.

Choosing an appropriate axis range involves a trade-off between participants' bias (over-reliance on the visual appearance of differences) and their sensitivity (capacity to identify actual differences) (Witt 2019). Just as a highly truncated y-axis can exaggerate trivial differences between values, an axis spanning the entire range of possible values can conceal important differences. Based on participants' judgements of effect size, Witt (2019) found that bias was reduced and sensitivity increased when using an axis range of approximately 1.5 standard deviations of the plotted data, compared to axes which spanned only the range of the data, or the full range of possible values. This provides further evidence of a powerful association between the appearance of data, when plotted, and subjective interpretations of differences between data points.

Further evidence of truncation effects, provided by Yang et al. (2021) improves on the design of previous studies which employed only a few observations per condition (Pandey et al. 2015) or small sample sizes (Witt 2019). Participants' ratings of the difference between two bars consistently provided evidence of the exaggerating effects of y-axis truncation. Yang et al. (2021) noted that increasing awareness does not eliminate this effect, which may function like an anchoring bias, in which numerical judgements are influenced by reference points (A. Tversky and Kahneman 1974). Another potential explanation discussed draws upon Grice's cooperative principle (Grice 1975). According to this account of effective communication, speakers are assumed to be in cooperation, and so will communicate in a manner that is informative, truthful, relevant, and straightforward. Analogously, a viewer will assume that a numerical difference in a chart must be genuinely large if it appears large, else it would not be presented that way. Effective visualisations should be designed so a viewer's instinctive characterisation of the data corresponds closely to their interpretation following a more detailed inspection (Yang et al. 2021).

Effects of Axis Limits on Magnitude Judgements

The above research consistently demonstrates that the magnitude of *the difference between values* is interpreted differently depending on the axis limits employed. The present investigation is concerned with how interpretations of the magnitude of *the values themselves* are affected by a chart's design.

Empirical evidence demonstrates that judgement of a value's magnitude can depend on its relationship to a grand total or to surrounding values. This can influence interpretation of verbal approximations, and also absolute values. For example, participants instructed to take 'a few' marbles picked up more when the total number available was larger (Borges and Sawyers 1974) and rated satisfaction with the same salary as higher when it appeared in the upper end of a range, compared to the lower end (Brown et al. 2008). As well as context, vertical position also plays a role in magnitude judgements. For example, children appear to intuitively

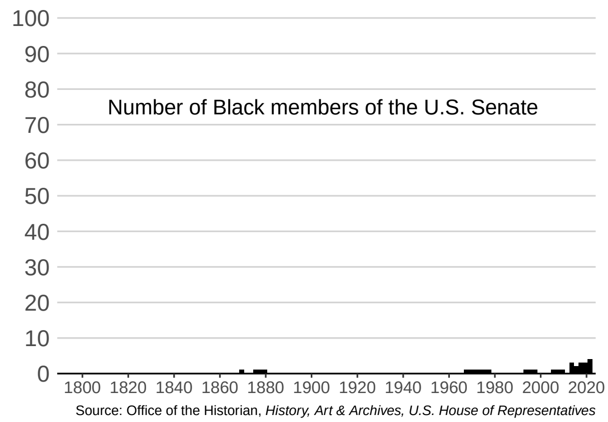

Figure 1: A reproduction of a bar chart from the New York Times. The y-axis limit is defined by the largest possible value, rather than the largest observed value, thus the magnitude of plotted values appears particularly small.

understand the relationship between height and value (Gattis 2002) Both the physical world, and language (e.g., spatial metaphors), provide countless examples where 'higher' is associated with 'more', and 'lower' with 'less', and this principle has been adopted as a convention in data visualisation (B. Tversky 1997).

In charts, inversions of the typical mapping between magnitude and vertical position charts can lead to misinterpretations (Okan et al. 2012; Pandey et al. 2015; Woodin, Winter, and Padilla 2022). Furthermore, when a company's financial performance was displayed entirely in the bottom fifth of a line chart, the company was perceived as less successful, compared to when the axis did not extend above the maximum value (Taylor and Anderson 1986). Sandman, Weinstein, and Miller (1994) investigated assessments of magnitude in risk ladders, where greater risks are presented at physically higher positions on a vertical scale. Participants rated asbestos exposure as a greater threat when it was plotted at a higher position, compared to a lower position.

The above findings can be regarded as preliminary evidence that changing axis limits may affect appraisals of data points' magnitudes. However, the evidence is not substantial. Taylor and Anderson (1986) did not disclose how judgements were elicited, or provide details of their sample size. Sandman, Weinstein, and Miller (1994) only explored responses to one specific risk (asbestos), and each participant only took part in a single trial. The perceived threat measure was a composite of several separate ratings, preventing diagnosis of whether manipulations affected interpretations of the plotted information in particular, or just related concepts. Further, both studies introduced a confounding variable by adjusting the difference between the minimum and maximum y-axis values across conditions. Stronger evidence is required regarding how axis limits may bias inferences about magnitude, and the cognitive mechanisms involved in generating these inferences.

Judgements of Event Outcomes

In the present study, participants viewed charts showing fictitious data on the chance of particular events occurring. This provided participants with a purpose; evaluating information about event outcomes is a more meaningful task than assessing how 'large' an abstract value is. Each value was represented using a single dot on a percentage probability scale. Our use of dot plots for conveying percentages was motivated by their simplicity and use of a single encoding channel (position), thus avoiding confounding variables from other encoding channels.

Presenting data about events with negative consequences warranted consideration of the cognitive processing of this information. These events are composed of two core components: 1) chance of occurrence and 2) outcome magnitude (severity). Individuals' assessments of chance and severity are not necessarily independent. Events are perceived as more likely when they are described as having more severe consequences (Harris and Corner 2011; Harris, Corner, and Hahn 2009). In a similar manner, events are associated with more substantial consequences when they are described as more likely (Kupor and Laurin 2020).

One account suggests that perceptions of probability and outcome magnitude are related because they are both assumed to reflect the potency of the event's cause (probability-outcome correspondence principle; (Keren and Teigen 2001)). According to this account, probabilities can occasionally provide meaningful indications of outcome magnitude (e.g., rainfall), but it is inappropriate to apply this perspective to all situations (e.g., volcanic eruptions). Therefore, even though charts in the present study only display the *chance* of events occurring, assessments of the *severity* of events' consequences may also differ between conditions. Collecting separate judgements for chance and severity of consequences provides a clearer picture of how the manipulation affects distinct aspects of participants' representations. Our use of Likert scales (with discrete options) rather than visual analogue scales (with continuous options; Sung and Wu (2018)) prevents participants from simply mapping probability percentages directly onto a linear scale.

Data Visualisation Literacy

When faced with charts that violate graphical conventions by using atypical scales, individuals with low data visualisation literacy are more likely to draw on data points' physical positions when making inferences about their magnitudes (Okan et al. 2012; Okan, Galesic, and Garcia-Retamero 2016). We administered Garcia-Retamero et al.'s (Garcia-Retamero et al. 2016) subjective graph literacy measure to determine whether responses to our manipulation of axis limits were associated with data visualisation literacy.

Experiments

We conducted three experiments manipulating y-axis limits in visualisations of fictitious data. This manipulation altered the physical positions of data points in a chart, but crucially the numerical values themselves remained the same.

Experiment 1 sought to establish whether y-axis limits affected magnitude judgements. To provide context for participants, text accompanying the charts outlined (fictitious) scenarios involving a specific negative outcome (e.g., loss on financial investment, delayed flights, etc.). Three plotted data points in each chart represented the chance of the negative outcome occurring (%) for three instances associated with the scenario (e.g., three investment opportunities, three airlines, etc.).

Experiment 2 introduced another factor in addition to the manipulation of y-axis limits. Half of the visualisations presented employed inverted y-axis orientations, where data points at lower physical positions represented greater values. This 2x2 experiment allowed us to investigate whether magnitude judgements were driven by data points' absolute positions, or their relative positions within the context of the axis limits.

Experiment 3 manipulated y-axis limits in inverted charts only, providing clarity on the ambiguous results of the previous experiment. Importantly, the use of inverted charts should not be considered an endorsement (see issues above). However, they serve to distinguish between two possible explanations, since they reverse the typical associations between physical position and magnitude.

Ethical approval was granted by The University of Manchester's Division of Neuroscience & Experimental Psychology Ethics Committee (Experiment 1: Ref 2021-11115-18258; Experiment 2: Ref 2021-11115-20464; Experiment 3: Ref. 2021-11115-20745). Data, analysis code, experimental scripts, materials and a link to run the experiments are available at https://osf.io/3epm2/. We also provide all necessary resources for running a Docker container, within which the computational environment used for analysis is recreated, meaning a fully-reproducible version of this paper can be generated.

Experiment 1

Method

Materials

Datasets

For each dataset, we generated three values from a normal distribution. Population means were specified manually in order to represent plausible values for the probability of the event occurring (28% - 72%). All datasets had a population standard deviation of 0.5. The same dataset was employed for both of the experimental conditions associated with a given event scenario.

Charts

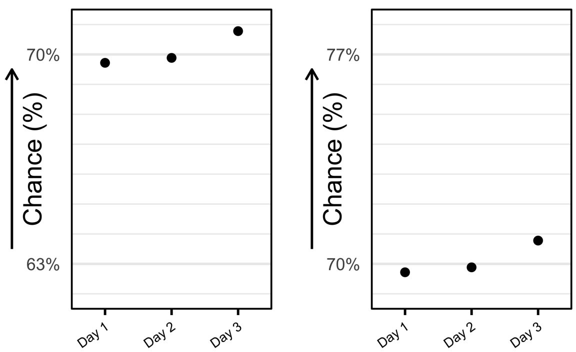

Figure 2: Example charts, taken from Experiment 1. The *high physical position* condition (left) presents data points near the top of the chart; the *low physical position* condition (right) presents the same data points near the bottom of the chart.

Datasets were displayed using dot plots. In experimental trials (n = 40), upper and lower axis limits were manipulated such that data points either appeared in the top third of the chart (high physical position: Figure 2 , left) or in the bottom third (low physical position: Figure 2, right).

The y-axis range in each chart was 10 percentage points. Horizontal gridlines appeared at one-unit increments. The horizontal gridlines 1.5 units from the extremes were labelled with numerical values.

Filler trials (n = 15) and attention check trials (n = 5) presented data points in the middle third of the chart. Filler trials employed this additional variation to prevent participants from identifying the purpose of the study.

Procedure

The experiment was programmed in PsychoPy (version 2021.1.4, (Peirce et al. 2019)) and hosted on pavlovia.org. Participants were instructed to complete the experiment on a desktop computer or laptop, not a tablet or mobile phone. Instructions explained that their task involved assessing the chance and severity of negative outcomes in various scenarios involving risks and noted that some scenarios might appear similar to other scenarios. Participants were asked to complete the task as quickly and accurately as possible. Two practice trials preceded the experiment proper.

An example of a single trial is shown in Figure 3. Participants provided two responses in each trial: a rating of the chance of the negative event occurring; and a rating of the severity of the consequences if that negative event occurred. Both 7-point Likert scales had two anchors at their extremes: *'Very unlikely'* and *'Very likely'*; for the 'Chance' scale and *'Very mild'* to *'Very severe'*. for the 'Severity' scale. All other points were unlabelled. Text specified that answers should be given in response to the plotted data (e.g., *"If you camp on one of these days…"*). The term 'chance' was used instead of 'probability' to avoid confusion with the standard 0-1 scale for probabilities, and to reflect casual usage.

Participants could change their responses as many times as they wished before proceeding to the next trial, but could not return to previous trials. In attention check trials, participants were instructed not to attend to the chart, and instead to provide specified responses on the Likert scales.

Before exiting the experiment, participants were informed that all presented data were fictitious and guidance was provided in case of distress.

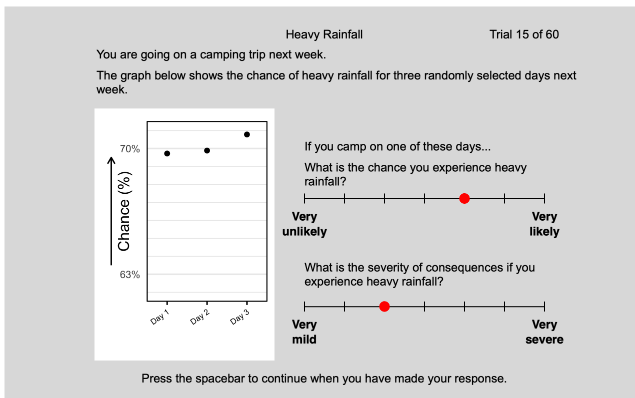

Figure 3: An example trial, taken from Experiment 1. Participants provided two ratings in each trial: the chance of an event occurring (magnitude rating), and the severity of consequences.

Design

We employed a repeated-measures, within-participants design. Participants encountered scenarios from experimental trials twice: once with data presented at a high physical position and once with data presented at a low physical position.

Materials were divided into two lists to minimise the likelihood of different versions of the same scenario appearing in close succession. One list contained half of the high-condition items and half of the low-condition items for the experimental scenarios. The other list contained the alternate versions of each of the experimental scenarios. Fillers and attention check questions were split between the two lists, and did not appear more than once. The order of the two lists was counterbalanced across participants. Within each list, scenarios were presented in a random order.

Participants

The experiments were advertised on Prolific.co, a platform for recruiting participants for online studies. Normal or corrected-to-normal vision and English fluency were required for participation.

Data were returned by 160 participants. Ten participants' submissions were rejected because they answered more than two of 10 attention check questions incorrectly. This left a total of 150 participants whose submissions were used for analysis (52.00% male, 45.33% female, 2.67% non-binary). Mean age was 31.49 ($SD = 12.47$)[1]. The mean data visualisation literacy score was 21.28 ($SD = 4.58$), out of a maximum of 30. Participants whose submissions were approved were paid £3.55. Average completion time was 25 minutes [2].

Analysis Technique

Analyses were conducted using R (version 4.2.1; (R Core Team 2022)).

Likert scales express granularity at the level of ordinal data. They record whether one rating is higher or lower than another, but not the magnitude of this difference. Therefore, Likert scales do not necessarily capture values from latent distributions (mental representations) in a linear manner. The distance between one pair of points and another pair may appear equal, but may represent different distances on the latent distribution. Therefore, it is inappropriate to analyse Likert scale data with metric models, such as linear regression (Liddell and Kruschke 2018). Throughout this paper, we construct cumulative link mixed-effects models, using the *ordinal* package (version 2019.12-10, (Christensen 2019)) to analyse Likert scale ratings. Odds ratio effect sizes were converted to Cohen' d values using the *effectsize* package (version 0.8.2, (Ben-Shachar, Lüdecke, and Makowski 2020)).

Selection of model random effects structures was automated using the *buildmer* package in R (version 2.3, (Voeten 2022)). The maximal random effects structure included random intercepts for participants and scenarios, plus corresponding slopes for the position variable (Barr et al. 2013). *buildmer* initially identified the most complex model which could successfully converge. It subsequently removed terms which did not contribute substantially to explaining variance in ratings.

Results

Magnitude Ratings

[1]Age data were unavailable for one participant.
[2]Timing data were unavailable for two participants.

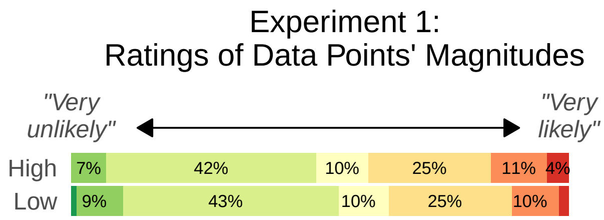

Figure 4: The distribution of Likert scale ratings of data points' magnitudes. The width of each response option represents the proportion of ratings recorded for that option. Note that data points presented at high physical positions (top) elicited a larger proportion of ratings on the right-hand side (representing greater magnitudes), compared to data points at low physical positions (bottom), which elicited a larger proportion of ratings on the left-hand side (representing smaller magnitudes).

Figure 4 plots the distribution of participants' ratings of data points' magnitudes, showing that values presented at high physical positions elicited a greater proportion of responses at the higher end of the rating scale than values presented at low physical positions.

A likelihood ratio test reveals that a model including physical position as a fixed effect explains significantly more variability in ratings than a model which does not include physical position as a fixed effect ($\chi^2(1) = 74.21$, $p < .001$). Data points' magnitudes were rated as greater when those data points were presented at high physical positions, compared to when the same data points were presented at low physical positions ($z = 8.57$, $p < .001$).

The odds ratio for the difference between conditions is 1.61 (95% CI [1.44, 1.79]). Participants were 1.6077133 times more likely to respond with a higher magnitude rating to data points presented at high positions than data points presented at low positions. This is equivalent to a Cohen's d value of 0.26.

This model included random intercepts for each participant and each scenario.

Severity Ratings

For ratings of the severity of consequences, a likelihood ratio test reveals that a model including physical position as a fixed effect explains significantly more variability in ratings than a model which does not include condition as a fixed effect: ($\chi^2(1) = 6.16$, $p = .013$). The severity of consequences was rated as greater when data points representing the chance of an event occurring were presented at high physical positions, compared to when the same data points were presented at low physical positions ($z = 2.50$, $p = .012$).

The odds ratio for the difference between conditions is 1.21, 95% CI [1.04, 1.41]. Participants were 1.2135597 times more likely to respond with a higher severity rating to data points presented at high positions than data points presented at low positions. This is equivalent to a Cohen's d value of 0.11.

This model employed random intercepts for each scenario, plus random intercepts and slopes for each participant. The slopes modelled, for each participant, the average difference between responses to data presented at different positions (henceforth referred to as *by-position slopes*).

Data Visualisation Literacy

We also generated two additional models, to test whether or not the above results could be explained by differences in data visualisation literacy. These models were identical to the above models except for the inclusion of participants' subjective data visualisation literacy scores as an additional fixed effect. Adjusting for participants' data visualisation literacy scores did not eliminate the effects of data points' positions on ratings of the magnitude of data points themselves ($z = 8.57$, $p < .001$, odds ratio = 1.61, 95% CI [1.44, 1.79]) or severity of consequences ($z = 2.51$, $p = .012$, odds ratio = 1.21, 95% CI [1.04, 1.41]).

Discussion

This experiment demonstrates that axis limits, which determine the position of plotted values, influence inferences about data points' magnitudes. Participants rated *the same values* as greater when these values were plotted at high positions, compared to low positions. Even though the charts only displayed data on the chance of negative outcomes occurring, ratings of severity of consequences were also greater when data points were presented at high positions. Accounting for differences in participants' data visualisation literacy did not alter the pattern of results.

Experiment 2

Introduction

Experiment 1 (E1) found that participants associated data points with greater magnitudes when those data points were positioned near the *top* of a chart, compared to when the same data points were positioned near the *bottom* of a chart.

One possible explanation for this finding is that participants made simple associations between absolute position and magnitude, equating physically higher data points with larger magnitudes and physically lower data points with smaller magnitudes. This is congruent with well-established conceptual metaphors for magnitude, where greater vertical positions denote greater magnitudes (B. Tversky 1997).

An alternative explanation is that participants used the y-axis as a frame of reference for assessing the magnitude of plotted values. For example, when considering data points near the bottom of the axis, participants may have recognised the potential for values larger than those observed, consequently associating plotted values with smaller magnitudes.

E1 does not provide a means of differentiating these competing explanations. Drawing inferences from data points' absolute positions would bias magnitude judgements in the same direction as drawing inferences from their relative positions. A high magnitude is implied by a data point's high physical position *and* its superior position in the context other of presented values. Therefore, further investigation is required in order to distinguish between the two competing explanations.

Plotting numerical values along the x-axis would not assist in answering this question, since values that are large in the context of the x-axis limits would be positioned on the right-hand side, which is also typically associated with larger magnitudes (Woodin and Winter 2018). However, inverting a vertical axis changes the typical relationship between physical position and numerical value: increasingly *lower* positions represent increasingly *higher* numerical values. This means data points presented near the *bottom* of a chart are numerically *larger* than the accompanying y-axis values. Therefore, inferences invoking relative numerical position would bias magnitude judgements in the opposite direction compared to inferences invoking data points' physical positions. This is illustrated in Figure 5.

In E2, we manipulate data points' physical positions by changing axis limits (as in E1), but *also* manipulate axis orientation, by employing conventional and inverted axes (in a 2 x 2 design). This allows us to identify the mechanism responsible for the previously observed bias in magnitude judgements. Use of absolute position would be indicated by higher magnitude ratings for data points at *high* physical positions (regardless of axis orientation). Alternatively, use of relative position would be indicated by higher magnitude ratings for data points at *high* physical positions in conventional charts and *low* physical positions in inverted charts.

Previous research suggests that charts with inverted axes can be prone to misinterpretation when viewers are not informed about the inversion (Pandey et al. 2015; Woodin, Winter, and Padilla 2022). Therefore, we provided explicit instruction to ensure participants were aware that inverted charts were presented.

Method

Materials

For this experiment, we used a Latin-squared design where participants only viewed one chart per scenario. In response to this, we increased the number of scenarios. This provided some compensation for the reduced experimental power caused by a reduction the number of observations per participant (as well as a reduction in participant numbers).

Two scenarios which were fillers in E1 were used as experimental scenarios[3] and three additional scenarios were created. One filler scenario was removed due to a concern about its quality (it concerned the risk to others as well as the risk to oneself). This resulted in a total of

[3]For one of these scenarios, the mean of the plotted data was also modified.

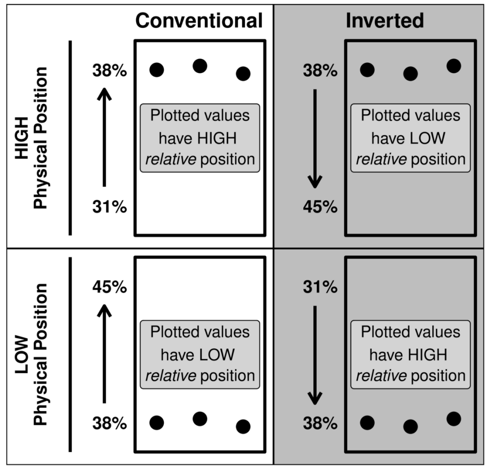

Figure 5: Rationale for Experiment 2: distinguishing the roles of absolute and relative position. In charts with conventional axis orientations (left column), there is congruence between data points' physical positions and their relative numerical positions in the chart. In charts with inverted axis orientations (right column), there is incongruence between data points' physical positions and their relative numerical positions in the chart. This allows us to test whether physical positions or relative numerical positions influence magnitude judgements.

24 experimental scenarios, 12 filler scenarios, and 5 attention check questions (41 trials in total).

Procedure

The experiment used PsychoPy version 2021.2.3. Participants specified the highest level of education they had received, in addition to answering demographic questions on age and gender. An additional slide in the instructions explained how to identify and interpret the different axis orientations, and encouraged participants to pay attention to this aspect of the charts:

> *You should pay attention to the direction of the arrow on the 'Chance' axis. If the arrow points upwards, the numbers in the graph get bigger as the axis goes up. Alternatively, if the arrow points downwards, the numbers get bigger as the axis goes down.*

Otherwise, the procedure was identical to E1.

Design

We employed a Latin-squared, within-participants design. Participants encountered each individual scenario only once, but were exposed to all combinations of physical plotting position and axis orientation throughout the experiment.

Participants

A viral social media post on 24th July 2021 endorsing the Prolific.co platform attracted many new users from a narrow demographic, skewing studies' participant distributions (Charalambides 2021). Therefore, the experiment was not advertised to users who signed-up to Prolific.co after 24th July 2021. The experiment was also not advertised to those who had participated in E1.

Data were returned by 129 participants. Per pre-registered exclusion criteria, five participants' submissions were rejected because they answered more than two of 10 attention check questions incorrectly. Submissions from four other participants were excluded from the final dataset for the following reasons: maximum completion time (67 minutes) was exceeded (two participants); the submission constituted second attempt following a saving error on first attempt (one participant); data were collected prior to pre-registration (one participant). This left a total of 120 participants whose submissions were used in the analysis (49.17% male, 50.83% female). Mean age was 30.73 ($SD = 17.83$). 100% had completed at least secondary education. The mean data visualisation literacy score was 21.72 ($SD = 4.70$), out of a maximum of 30. Participants whose submissions were approved were paid £3.55. Average completion time was 21 minutes.

Results

Magnitude Ratings

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal." *Journal of Memory and Language* 68 (3): 255–78. https://doi.org/10.1016/j.jml.2012.11.001.

Ben-Shachar, Mattan, Daniel Lüdecke, and Dominique Makowski. 2020. "Effectsize: Estimation of Effect Size Indices and Standardized Parameters." *Journal of Open Source Software* 5 (56): 2815. https://doi.org/10.21105/joss.02815.

Borges, Marilyn A., and Barbara K. Sawyers. 1974. "Common Verbal Quantifiers: Usage and Interpretation." *Journal of Experimental Psychology* 102 (2): 335–38. https://doi.org/10.1037/h0036023.

Brown, Gordon D. A., Jonathan Gardner, Andrew J. Oswald, and Jing Qian. 2008. "Does Wage Rank Affect Employees' Well-Being?" *Industrial Relations* 47 (3): 355–89. https://doi.org/10.1111/j.1468-232X.2008.00525.x.

Charalambides, Nick. 2021. "We Recently Went Viral on TikTok - Here's What We Learned." https://blog.prolific.co/we-recently-went-viral-on-tiktok-heres-what-we-learned/.

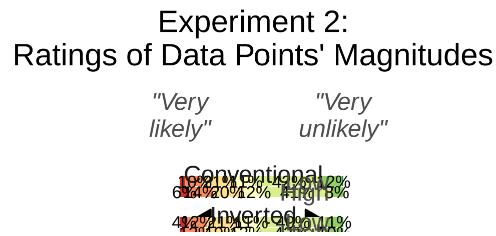

Figure 6: Participants rated the chance of each negative event occurring on a 7-point Likert scale. The distribution of ratings, ranging from "Very unlikely" (far left, dark green) to "Very likely" (far right, red) is shown separately for each combination of the levels of each condition (axis orientation: conventional, inverted; data points' physical position: high, low). Note that the pattern of responses to data presented at different positions in the Conventional Axis condition appears to be the opposite to the pattern for Inverted Axis condition. When charts used conventional axes, greater magnitude ratings were more common for data presented at high physical positions, whereas when charts used inverted axes, greater magnitude ratings were more common for data presented at low physical positions.

Christensen, Rune Haubo B. 2019. "Ordinal—Regression Models for Ordinal Data." https://CRAN.R-project.org/package=ordinal.

Cleveland, W. S., P. Diaconis, and R. McGill. 1982. "Variables on Scatterplots Look More Highly Correlated When the Scales Are Increased." *Science* 216 (4550): 1138–41. https://doi.org/10.1126/science.216.4550.1138.

Correll, Michael, Enrico Bertini, and Steven Franconeri. 2020. "Truncating the Y-Axis: Threat or Menace?" In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. Honolulu HI USA: ACM. https://doi.org/10.1145/3313831.3376222.

Garcia-Retamero, Rocio, Edward T. Cokely, Saima Ghazal, and Alexander Joeris. 2016. "Measuring Graph Literacy Without a Test: A Brief Subjective Assessment." *Medical Decision Making* 36 (7): 854–67. https://doi.org/10.1177/0272989X16655334.

Gattis, Merideth. 2002. "Structure Mapping in Spatial Reasoning." *Cognitive Development* 17 (2): 1157–83. https://doi.org/10.1016/S0885-2014(02)00095-3.

Grice, Paul. 1975. "Logic and Conversation." In *Syntax and Semantics Vol.3: Speech Acts*, edited by P Cole and J. L. Morgan, 41–58. New York: Academic Press.

Harris, Adam J. L., and Adam Corner. 2011. "Communicating Environmental Risks: Clarifying the Severity Effect in Interpretations of Verbal Probability Expressions." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37 (6): 1571–78. https://doi.org/10.1037/a0024195.

Harris, Adam J. L., Adam Corner, and Ulrike Hahn. 2009. "Estimating the Probability of Negative Events." *Cognition* 110 (1): 51–64. https://doi.org/10.1016/j.cognition.2008.10.006.

Keren, Gideon, and Karl Halvor Teigen. 2001. "The Probability—Outcome Correspondence

Principle: A Dispositional View of the Interpretation of Probability Statements." *Memory and Cognition* 29 (7): 1010–21. https://doi.org/10.3758/BF03195763.

Kupor, Daniella, and Kristin Laurin. 2020. "Probable Cause: The Influence of Prior Probabilities on Forecasts and Perceptions of Magnitude." Edited by Chris Janiszewski and J Jeffrey Inman. *Journal of Consumer Research* 46 (5): 833–52. https://doi.org/10.1093/jcr/ucz025.

Liddell, Torrin M., and John K. Kruschke. 2018. "Analyzing Ordinal Data with Metric Models: What Could Possibly Go Wrong?" *Journal of Experimental Social Psychology* 79 (November): 328–48. https://doi.org/10.1016/j.jesp.2018.08.009.

Okan, Yasmina, Mirta Galesic, and Rocio Garcia-Retamero. 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): 271–94. https://doi.org/10.1002/bdm.1891.

Okan, Yasmina, Rocio Garcia-Retamero, Mirta Galesic, and Edward T. Cokely. 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, Anshul Vikram, Katharina Rall, Margaret L. Satterthwaite, Oded Nov, and Enrico Bertini. 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 - CHI '15*, 1469–78. Seoul, Republic of Korea: ACM Press. https://doi.org/10.1145/2702123.2702608.

Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. "PsychoPy2: Experiments in Behavior Made Easy." *Behavior Research Methods* 51 (1): 195–203. https://doi.org/10.3758/s13428-018-01193-y.

R Core Team. 2022. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sandman, Peter M., Neil D. Weinstein, and Paul Miller. 1994. "High Risk or Low: How Location on a "Risk Ladder" Affects Perceived Risk." *Risk Analysis* 14 (1): 35–45. https://doi.org/10.1111/j.1539-6924.1994.tb00026.x.

Sung, Yao-Ting, and Jeng-Shin Wu. 2018. "The Visual Analogue Scale for Rating, Ranking and Paired-Comparison (VAS-RRP): A New Technique for Psychological Measurement." *Behavior Research Methods* 50 (4): 1694–1715. https://doi.org/10.3758/s13428-018-1041-8.

Taylor, Barbara G, and Lane K Anderson. 1986. "Misleading Graphs: Guidelines for the Accountant." *Journal of Accountancy* 162 (4): 126–35.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31. https://doi.org/10.1126/science.185.4157.1124.

Tversky, Barbara. 1997. "Cognitive Principles of Graphic Displays." In *Proceedings of the AAAI 1997 Fall Symposium on Reasoning with Diagrammatic Representations*, 116–24. Menlo Park, CA: AAAI Press,.

Voeten, Cesko C. 2022. "Buildmer: Stepwise Elimination and Term Reordering for Mixed-

Effects." https://CRAN.R-project.org/package=buildmer.

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

Woodin, Greg, and Bodo Winter. 2018. "Placing Abstract Concepts in Space: Quantity, Time and Emotional Valence." *Frontiers in Psychology* 9 (2169): 1–14. https://doi.org/10.3389/fpsyg.2018.02169.

Woodin, Greg, Bodo Winter, and Lace Padilla. 2022. "Conceptual Metaphor and Graphical Convention Influence the Interpretation of Line Graphs." *IEEE Transactions on Visualization and Computer Graphics* 28 (2): 1209–21. https://doi.org/10.1109/TVCG.2021.3088343.

Yang, Brenda W., Camila Vargas Restrepo, Matthew L. Stanley, and Elizabeth J. Marsh. 2021. "Truncating Bar Graphs Persistently Misleads Viewers." *Journal of Applied Research in Memory and Cognition* 10 (2): 298–311. https://doi.org/10.1016/j.jarmac.2020.10.002.