Absence Makes The Chart Grow Stronger: Blank Space and Axis Range Influence Interpretations of Magnitude in Risk Communication

Duncan Bradley, Gabriel Strain, Caroline Jay, Andrew J. Stewart

Abstract—When visualizing data, chart designers have the freedom to choose the upper and lower limits of their numerical axes. Axis limits determine the physical positions of plotted values, and can introduce substantial blank space. For charts presenting data on the chance of negative events occuring, manipulating axis limits affects viewers' interpretations of plotted values' magnitudes, influencing understanding of the risk information being communicated. Across three experiments (total N=420), we demonstrate that, surprisingly, participants did not simply equate values presented at higher vertical positions with greater magnitudes. Instead, they used the numerical context supplied by axis limits to assess the magnitude of data points by contrasting these values against accompanying blank space. Data points were considered larger when they were numerically greater than the plausible values implied by blank space, *even* when they were presented at the *bottom* of a chart. Chart designers must consider the role of their axis range in viewers' interpretations of the magnitudes of plotted data points. We recommend displaying the range of relevant values in order to communicate the specific context for each dataset.

Index Terms—Cognition, framing effects, chart design, axis range, magnitude judgements.

1 Introduction

Context is crucial for effectively judging the magnitude of numbers. A 40% probability is twice as great as a 20% probability, but in the absence of context, it is unclear whether this value should be considered large or small. For the chance of experiencing post-surgery complications, 40% may be considered large, but may be considered small for the chance that a laboratory test can detect a disease.

In charts, numerical axes often provide contextual cues for judging the magnitude of plotted values. The range of values on an axis provides a frame of reference for assessing whether a data point is numerically large or small. Figure 1 (a reproduction of a bar chart from the New York Times), which plots over time the number of Black members of the U.S. senate [15], provides a striking illustration. Unusually, the continuous y-axis does not terminate just above the highest plotted value. Instead, it 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 highlights the absent data points: the vast majority of senators who are not Black. The visual arrangement communicates the magnitude of the plotted values in context.

It is unclear exactly how an axis range influences a viewer's inferences about magnitude. One possible explanation is that the unfilled area indicates the range of plausible values. That is, plotted values may be judged as small in magnitude because the potential for substantially larger values is clearly displayed. Alternatively, viewers' assessments may be influenced by the appearance of plotted values only, and not by contrast with blank space. Viewers may simply interpret the magnitude of data points at higher positions as 'high' and those at lower position as 'low', ignoring the plausible alternative values implied by blank space. The present set of experiments explores which of these two accounts explains how axis ranges contribute to the communication of magnitude.

1.1 Effects of Context on Magnitude Judgments

Empirical evidence demonstrates that judgment 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 ([2]) and rated satisfaction with the same salary as higher when it appeared in the upper end of a range, compared to the lower end [3].

1.2 Effects of Axis Limits on Comparison of Values

Several studies have explored how axis limits can alter impressions of the *relationships between* presented values, rather than the magnitudes of values themselves. When axis ranges are expanded to create blank space around a cluster of data points, correlation between those points is judged as stronger [6]. In bar charts, participants rate the differences between values as greater when the vertical gap between bars is larger, due to a truncated y-axis [25]. Correll et al.'s [7] experiments found that greater truncation resulted in higher effect-size judgments in both line charts and bar charts. Truncation effects persisted even when participants estimated the values of specific data points, suggesting this bias is driven by initial impressions, rather than a misinterpretation of the values portrayed by graphical markings. Correll et al. [7] found no reduction in effect size judgments when truncation was communicated using graphical techniques (e.g., axis breaks and gradients). The unavoidable consequence, they 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 visually recognize actual differences). 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 [35]. Based on participants' judgments of effect size, Witt [35] 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. [37] improves on the design of previous studies which employed only a

- *Duncan Bradley is with the Division of Neuroscience and Experiment Psychology, The University of Manchester, UK. Email: duncan.bradley@manchester.ac.uk.*
- *Gabriel Strain and Caroline Jay are with the Department of Computer Science, The University of Manchester, UK. Email: {gabriel.strain | caroline.jay}@manchester.ac.uk.*
- *Andrew J. Stewart is with the Division of Neuroscience and Experiment Psychology and the Department of Computer Science, The University of Manchester, UK. Email: andrew.stewart@manchester.ac.uk.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

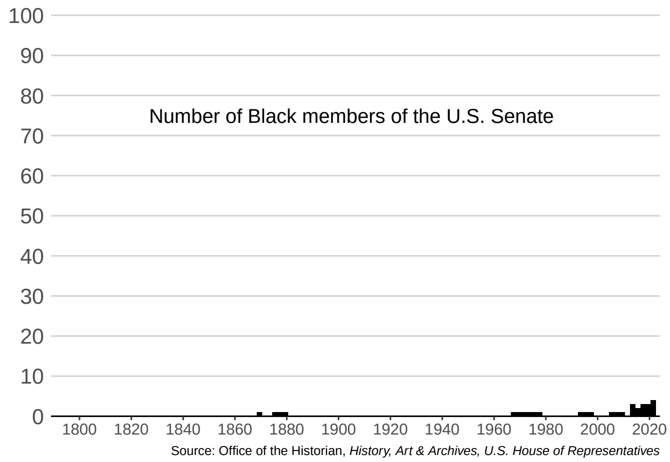

Fig. 1. In this chart, the y-axis limit is the largest possible value, rather than the largest observed value, so plotted values appear to have particularly small magnitudes.

few observations per condition [25] or very small sample sizes [35]. Participants' ratings of the difference between two bars consistently provided evidence of the exaggerating effects of y-axis truncation. Yang et al. [37] noted that increasing awareness does not eliminate the effect, which may function like an anchoring bias, where numerical judgments are influenced by reference points [32]. Another potential explanation discussed draws upon Grice's cooperative principle [12]. 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 visualizations should be designed so a viewer's instinctive characterization of the data corresponds closely to their interpretation following a more detailed inspection [37].

1.3 Effects of Axis Limits on Extraction of Values

The above research consistently demonstrates that the magnitude of *the difference between values* is interpreted differently depending on the appearance of the data points when plotted. The present investigation is concerned with how interpretations of the magnitude of *the values themselves* are affected by their visual properties. From a cognitive processing perspective, vertical position is a strong indicator of magnitude. For example, children appear to intuitively understand the relationship between height and value [11]. 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 visualization [33].

Research on data visualizations has identified cases where the relationship between magnitude and vertical position can influence interpretation. For example, inversions of this mapping in charts can lead to misinterpretations [22, 25, 36]. 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 than when no blank space appeared above the maximum value [31]. Sandman et al. [29] investigated assessments of magnitude in risk ladders, where greater risks are presented at physically higher positions on a vertical scale. Participants rated the threat of asbestos exposure higher when it was plotted at a higher 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 [31] did not disclose how judgments were elicited, or provide details of their sample size. Sandman et al. [29] only explored responses to one specific risk (asbestos), and each participant only took part in a single trial. In addition, the 'threat' 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. To understand how different displays of the same values elicit different inferences about magnitude, and to provide recommendations for best practice, stronger evidence is required, as is investigation into the cognitive mechanisms involved in generating these inferences.

1.4 The Present Experiments

In a set of three experiments employing a large number of observations, we investigate how employing different axis limits affects interpretations of the magnitude of plotted values. This manipulation changes the context surrounding data points, and their physical positions, but crucially the numerical values themselves remain the same.

All data visualizations used in the present set of experiments displayed the chance of negative events occurring. This provides participants with a purpose in the experiments; evaluating information in such risk scenarios is a more meaningful task than assessing, in an abstract manner, how 'large' a value is. Furthermore, charts are frequently used for the communication of such risks, and manipulating aspects of a chart can change interpretations of the risks displayed [8, 9, 17, 24, 38].

Risk 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. An event is perceived as more likely when it is described as having more severe consequences [13, 14]. In a similar manner, an event is associated with more substantial consequences when it is described as more likely [19]. 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; [18]. 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 set of experiments only display the chance of events occurring, assessments of the severity of events' consequences may also differ between conditions. Collecting separate judgments of chance and severity of consequences for each scenario provides a clearer picture of how the manipulation affects distinct aspects of participants' representations of risk. Use of Likert scales (with discrete options) rather than visual analogue scales (with continuous options; [30]) prevents participants from simply mapping probability percentages directly onto a linear scale. We also administered a subjective graph literacy measure, to determine the degree to which our manipulation(s) affect interpretations after accounting for differences in graph literacy. Previous research has shown that responses to visualizations which violate graphical conventions by using atypical scales suggest individuals with lower graph literacy are more likely to draw on data points' physical positions when making inferences about their magnitudes [23, 22].

1.4.1 Open Research Statement

All experiments in this paper were pre-registered (`https://osf.io/qn46s/`). There are no diversions from pre-registered experimental designs, exclusion criteria or sample size. However, the reported analyses differ in some respects from the pre-registered protocol. For full transparency, we outline these diversions here.

Consistent with our pre-registration, when building models for our main analyses, we sought the most complex random effects structures that would successfully converge. These model structures were identified by the *buildmer* package in R [34], which subsequently removed terms which did not contribute substantially to explaining variance in ratings. This means that the final model used in analysis was not always the most complex converging model.

In pre-registrations for Experiment 2 and Experiment 3, we proposed testing for an interaction between our manipulation(s) and graph literacy. However, this was motivated by a concern about whether accounting for graph literacy could explain the presence or absence of

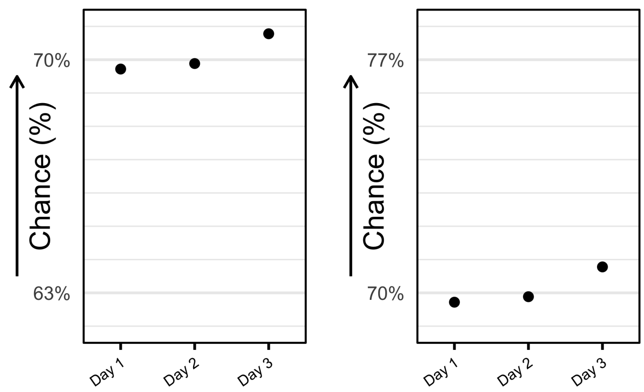

Fig. 2. Example Charts. The 'high physical position' condition (left) presents data points near the top of the chart; the 'low physical position' condition (right) presents the same data points near the bottom of the chart.

effects of our manipulation(s). Therefore, we substitute these planned analyses with a more appropriate approach, treating graph literacy as a co-variate only (no interaction). This matches the pre-registered analysis from Experiment 1, providing consistency across the three experiments. Due to this revision, pre-registered hypotheses about graph literacy are not discussed.

2 Experiment 1

2.1 Introduction

Our initial experiment investigated whether changing axis limits affects interpretation of data points' magnitudes. For the different versions of each chart, we presented the same data points at different vertical positions by altering both the upper and lower y-axis limits.

We predicted that ratings of data points' magnitudes (chance of occurrence) and/or ratings of the severity of consequences would be greater when data points were presented at higher physical positions, compared to when the same data points were presented at lower positions.

2.2 Methods

2.2.1 Materials

Text and an accompanying chart were presented in each trial. Two sentences outlined a scenario involving a risk, and explained what the chart depicted. For example:

> *You are going on a camping trip next week. The graph below shows the chance of heavy rainfall for three randomly selected days next week.*

The accompanying dot plot displayed the chance (as a percentage) of a negative outcome occurring, for three options associated with the scenario (Figure 2). The label 'Chance' was used instead of 'Probability' to avoid confusion with the standard 0-1 scale for probabilities, and to reflect casual usage.

In experimental trials (n = 40), all three data points were either plotted in the top third of the chart (high physical position: Figure 2, left) or in the bottom third of the chart (low physical position: Figure 2, right). The plotted dataset differed for each distinct scenario, but was identical for the two charts associated with a given scenario. In filler trials (n = 15) and attention check trials (n = 5), data points were plotted in the middle third of the chart.

The y-axis range in each chart was 10 percentage points. Horizontal gridlines appeared at one-unit increments. In all trials, the gridline 1.5 percentage points above the bottom of the chart was labelled with a numerical value, as was the gridline 1.5 percentage points below the top of the chart.

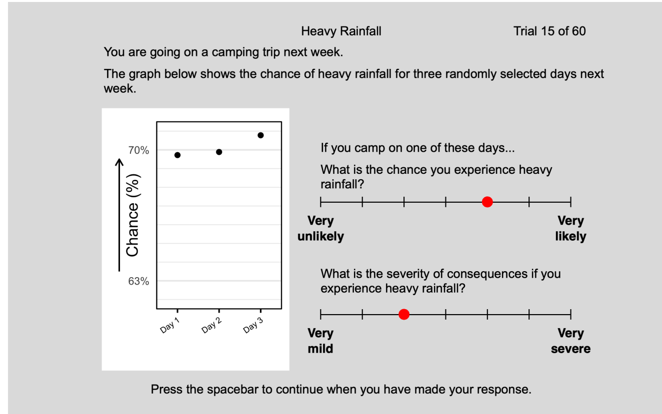

Fig. 3. Example Trial. Participants rated the chance and severity of negative outcomes in each trial.

2.2.2 Procedure

The experiment was programmed in PsychoPy (version 2021.1.4; [26]) and hosted on pavlovia.org. Participants were instructed to complete the experiment on a desktop computer or laptop, not a tablet or mobile phone. After providing informed consent, participants submitted their age and gender, and completed a five-item subjective graph literacy scale [10]. They were reminded that the experiment involved information about risks, and could cause distress, so were entitled to withdraw from the experiment at any time. Following this, instructions explained that their task involved assessing the chance and severity of negative outcomes in various scenarios involving risks. The instructions 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 were presented before the experiment proper began.

Two responses were required for 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. Above these Likert scales, a short phrase indicated that the questions should be answered in response to the plotted data (e.g., *"If you camp on one of these days..."*).

Each Likert scale had two anchors at its extremes, but all other points were unlabeled. The leftmost option in the 'chance' Likert scale was *'Very unlikely'*, and the rightmost option *'Very likely'*. The leftmost option in the 'severity' Likert scale was *'Very mild'* and the rightmost option *'Very severe'*. Likert scales appeared on the same screen as the text and chart (Figure 3). Participants were permitted to change their responses as many times as they wished before proceeding to the next trial, but could not return to previous trials.

Attention check trials (n = 5) followed the same layout, with text, a chart, and Likert scales, but the task differed. Participants were instructed not to attend to the chart, and instead to provide specified responses on the Likert scales. For example:

> *You are expected to stay on task throughout this experiment. For this trial, ignore the graph below. Respond 'Very unlikely' on the top scale, and 'Very mild' on the bottom scale.*

For attention check trials, the questions above the Likert scales were *"What is the chance response specified above?"* and *"What is the severity response specified above?"*.

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

2.2.3 Design

We employed a repeated-measures, within-participants design. In experimental trials, participants encountered each scenario twice: once with data presented at a high physical position and once with data presented at a low physical position. In each trial, participants rated the chance of a negative event occurring, and the severity of its consequences, on seven-point Likert scales.

Materials were divided into two lists to minimize the likelihood of different versions of the same scenario appearing in close succession. In one list, half of the experimental scenarios were accompanied by charts displaying data at high physical positions, and half were accompanied by charts displaying data at low physical positions. 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, and within each list, scenarios were presented in a random order.

2.2.4 Participants

The experiment was advertised on Prolific.co, a platform for recruiting participants for online studies. A viral social media post on 24th July 2021 endorsing the website attracted many new users from a narrow demographic, skewing studies' participant distributions [4], however, data for this experiment were collected prior to this. Normal or corrected-to-normal vision and English fluency were required for participation.

Data were returned by 160 participants. Per pre-registered exclusion criteria, 10 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 graph literacy score was 21.28 (*SD* = 4.58), out of a maximum of 30. Participants whose submissions were approved were paid £3.55, and average completion time was 25 minutes [2]. Ethical approval was granted by The University of Manchester's Division of Neuroscience & Experimental Psychology Ethics Committee (Ref. 2021-11115-18258).

2.3 Analysis

Analyses were conducted using R (version 4.1.2, [27]). Raw data and analysis scripts are available at `https://github.com/duncanbradley/position_magnitude`.

Likert scales only express granularity at the level of ordinal data. They record whether one rating is higher or lower than than another, but do not record the magnitude of this difference. Therefore, Likert scales do not capture values from latent distributions (mental representation) in a linear manner. On a Likert scale, the distance between one pair of points and another pair may appear equal, but may represent very different distances on the latent distribution. Therefore, it is inappropriate to analyse Likert scale data with metric models, such as linear regression [21]. Throughout this paper, we construct cumulative link mixed-effects models, using the *ordinal* package (version 2019.12-10, [5]) to analyse Likert scale ratings.

Selection of random effects structures for models was automated using the *buildmer* package (version 2.3, [34]). The maximal random effects structure included random intercepts for participants and scenarios, plus corresponding slopes for fixed effects terms [1]. From this formula, *buildmer* initially identified the most complex model which could successfully converge, prioritizing the terms which explained the most variance in the data, then eliminated terms which did not provide significant contributions (assessed using likelihood ratio tests).

Figure 4 plots the distribution of participants' ratings of data points' magnitudes, for data points presented at high and 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 ($z = 8.57$, $p < .001$). This model employed random intercepts for each scenario and each participant. Estimated marginal means, calculated using the *emmeans* package (version 1.7.0, [20] for these ratings are plotted in Figure **?**](fig:r1-c-emm-plot).

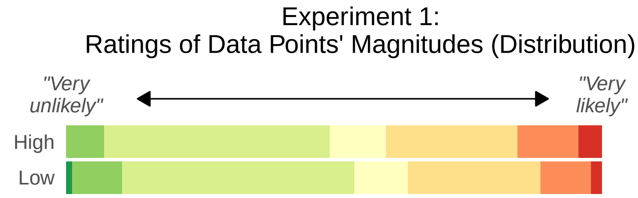

Fig. 4. Participants rated the chance of each negative event occurring on a 7-point Likert scale. The distribution of ratings, ranging from "Very unlikely" (far left, dark green) to "Very likely" (far right, red), is shown separately for charts where values were presented at a high physical position (top) and a low physical position (bottom). Note that data points at high physical positions elicited a larger proportion of ratings on the right-hand side (which represents greater magnitudes), compared to data points at low physical positions, which elicited a larger proportion of ratings on the left-hand side (representing smaller magnitudes).

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 ($z = 2.50$, $p = .012$). This model employed random intercepts for each scenario, plus random intercepts and slopes for each participant. The slopes modeled, for each participant, the average difference between responses to data presented at different positions (henceforth referred to as 'by-position slopes').

We also generate two additional models, to test whether or not the above results could be explained by differences in graph literacy. These models were identical to the above models except for the inclusion of participants' graph literacy scores as an additional fixed effect. Adjusting for participants' graph literacy scores did not eliminate the effects of data points' positions on ratings of the magnitude of data points themselves ($z = 8.57$, $p < .001$) or severity of consequences ($z = 2.51$, $p = .012$).

The above analysis employs models with *flexible* thresholds. This allows for variable distances between decision thresholds in the models (points on the latent distribution dividing responses between two categories). Comparison with models that specify *equidistant* thresholds reveals that models with flexible thresholds are superior, for ratings of the magnitude of data points themselves ($\chi^2(4) = 609.44$, $p < .001$) and ratings of the severity of consequences ($\chi^2(4) = 142.84$, $p < .001$. flexible). This suggests participants treated intervals between response categories as irregular, and validates the use of flexible thresholds in model construction.

2.4 Discussion

Participants rated the magnitudes of data points as greater when those data points were presented near the top of the chart, compared to when the same data points were presented near the bottom.

Higher bars and ascending lines typically represent higher numbers and ascending trends, so within a single chart, inferring that values presented higher up are greater than those lower down will often be correct in normal usage. This experiment, however, establishes that inferences about the magnitude of *the same value* can change depending on its position. Modeling differences in participants' graph literacy did not remove the influence of our experimental manipulation on interpretations.

[1]Age data was unavailable for one participant, but was available for all other participants in the dataset.

[2]Timing data was unavailable for two participants, but was available for all other participants in the dataset.

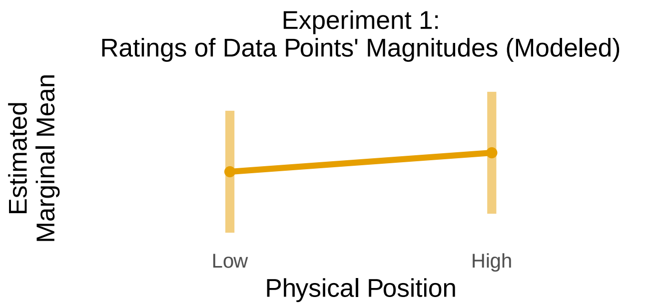

Fig. 5. Estimated marginal means for ratings of data points' magnitudes (generated by the cumulative-link mixed model). Magnitudes were rated as greater when data points were presented at high physical positions. Translucent bars show 95% confidence intervals.

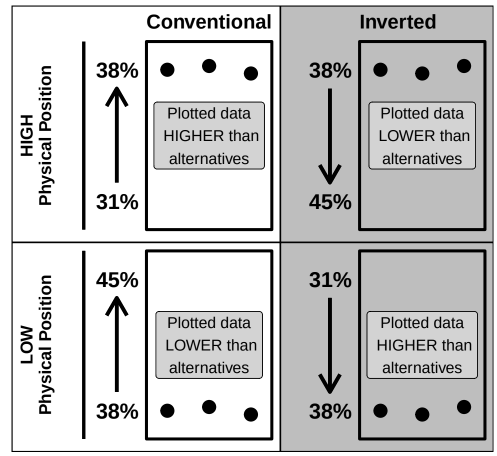

Fig. 6. Rationale for Experiment 2: Distinguishing the Roles of Absolute and Relative Position. In charts with conventional axis orientations (left column), there is congruity between data points' absolute positions and their relative positions in the chart. In charts with inverted axis orientations (right column), there is incongruity between data points' absolute positions and their relative positions in the chart. For example, at high absolute positions in conventional charts (top left), data points are relatively higher than implied alternatives. But at the same absolute positions, in inverted charts, the same values are relatively lower than alternatives (top right).

Ratings of both the chance of an event occurring, and the severity of consequences, were affected by the manipulation of axis limits and data points' positions, even though the charts only displayed data on the former. This accords with previous reports of an interplay between properties of presented information and impressions of related but distinct concepts, in particular the finding that higher prior probabilities were associated with impressions of greater event magnitudes [19]. However, it is unclear whether the effects of different axis ranges on interpretations of magnitude are driven by an association between a data point's *absolute* position and its magnitude, or an association between its *relative* position and its magnitude. If absolute position influences interpretations, mentally representing the magnitude of a data point may simply involve associating data points at higher positions with higher values (and lower positions with lower values). In contrast, if relative position influences interpretations, mentally representing the magnitude of a data point would involve a comparison with plausible alternative values, which are not plotted, but implied through use of blank space. This important distinction is explored in Experiment 2.

3 Experiment 2

3.1 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 and substantial blank space appeared *below* them, compared to when the same data points were positioned near the *bottom* of a chart, with substantial blank space *above*.

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 relates to well-established conceptual metaphors for magnitude, where greater vertical positions denote greater magnitudes [33].

An alternative explanation is that participants used blank space as a reference point when assessing the magnitude of plotted values. For example, when viewing substantial blank space above plotted data points, participants may have recognized the potential for values larger than those observed, consequently associating plotted data points with smaller magnitudes.

E1 does not provide a means of differentiating these competing explanations. Drawing inferences from data points' absolute positions would orient magnitude judgments in the same direction as drawing inferences from their positions relative to blank space. A high magnitude is implied by a data point's high physical position *and* the presence of substantial blank space below. Therefore, an additional experiment is required in order to distinguish between the two competing explanations.

Inverting a vertical axis changes the 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, with substantial blank space above, are numerically *larger* than the plausible values represented by this blank space. This is illustrated in Figure 6. Therefore, inferences invoking blank space would generate the opposite impressions to inferences invoking data points' physical positions only.

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). If interpretations of magnitude differ according to whether data points are smaller or larger than other plausible values implied by the chart (regardless of physical position), this will demonstrate that interpretations are driven by positions relative to blank space, rather by absolute position.

Previous research suggests that charts with inverted axes can be prone to misinterpretation when viewers are not informed about the inversion (Pandey et al. [25]; Woodin et al. [36]). In E2, we provide explicit instruction to ensure participants are aware that inverted charts are presented.

For charts with conventional axis orientations, we predicted in our pre-registration that results from E1 would be replicated. That is, data points presented at higher physical positions would be associated with greater magnitude ratings, compared to data points presented at lower physical positions. For charts with inverted axis orientations, we outlined what different patterns of magnitude ratings would signal about the mechanism used to interpret magnitude. Specifically, use of absolute position would be indicated by greater magnitude ratings for data points at *higher* physical positions (and therefore no difference compared to conventional charts). Alternatively, use of position relative to blank space would be indicated by greater magnitude ratings for data points at *lower* physical positions (and therefore the opposite pattern compared to conventional charts).

3.2 Method

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

3.2.2 Procedure

The experiment was programmed in PsychoPy (version 2021.2.3 Peirce et al. [26]). 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:

> *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.

3.2.3 Design

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

3.2.4 Participants

The experiment was not advertised on Prolific.co to those who had participated in E1, or those who signed-up to Prolific.co after 24th July 2021 (due to the shift in participant demographics). Normal or corrected-to-normal vision and English fluency were required for participation.

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 29.32 (*SD* = 10.45). 100% had completed at least secondary education. The mean graph literacy score was 21.73 (*SD* = 4.70). Participants whose submissions were approved were paid £2.37, and average completion time was 21 minutes. Ethical approval was granted by The University of Manchester's Division of Neuroscience & Experimental Psychology Ethics Committee (Ref. 2021-11115-20464).

3.3 Analysis

Figure 7 plots the distribution of participants' ratings of data points' magnitudes, for data points presented at high and low physical positions, in charts with conventional axis orientations and inverted axis orientations. A likelihood ratio test reveals that a model including the interaction between physical position and axis orientation as a fixed effect explains significantly more variability in ratings than a model without this interaction as a fixed effect ($\chi^2(1) = 8.22$, p = .004). There was a significant interaction between physical position and y-axis orientation (z = 2.91, p = .004). This interaction is plotted in Figure 8. This model employed random intercepts and by-position slopes for each scenario. Random intercepts were included for each participant, as well as slopes capturing differences in participants' responses to data presented at different positions, different orientations, and the interaction between these.

Pairwise comparisons (with Sidak adjustment) reveal that the effect of position in charts with conventional y-axis orientations (E1) was replicated (z = 3.56, p = .001). Data points' magnitudes were rated as greater when they were presented at high physical positions, compared to when they were presented at low physical positions. There was no significant difference between magnitude ratings for data points plotted at different positions when inverted axes were used (z = -1.39, p = .512). Therefore, we observe a different pattern of results when an inverted axis is used, compared to when a conventional axis is used. This suggests that differences in ratings for data points at different positions in physical space are not due to simple associations between vertical position and magnitude. The interaction remained when controlling for graph literacy: z = 2.91, p = .004, and when controlling for list number: z = 2.92, p = .004.

For ratings of the severity of consequences, a likelihood ratio test reveals that a model including the interaction between physical position and axis orientation as a fixed effect explains significantly more variability in ratings than a model without this interaction as a fixed effect ($\chi^2(1) = 5.13$, p = .024).There was a significant interaction between physical position and y-axis orientation (z = 2.28, p = .022). This model employed random intercepts for each scenario. Random intercepts were included for each participant, as well as slopes capturing differences in participants' responses to data presented at different positions, different orientations, and the interaction between these. Despite the interaction, the main effect in severity ratings from E1, different responses to data points at different positions in conventional charts, was not replicated (1.53, p = .414). There was also no evidence of different responses to data points at different positions in inverted charts (-1.54, p = .412). This interaction appears to be driven by a weak and likely spurious difference between ratings for data points at high physical positions in inverted and conventional charts (-2.52, p = .047). The interaction remained when controlling for graph literacy: z = 2.29, p = .022, and when controlling for list number: z = 2.28, p = .023.

Models employing flexible decision thresholds (as above) were superior to models employing equidistant thresholds, for ratings of the magnitude of data points themselves ($\chi^2(4) = 346.93$, $p < .001$), and ratings of the severity of consequences: ($\chi^2(4) = 74.10$, $p < .001$).

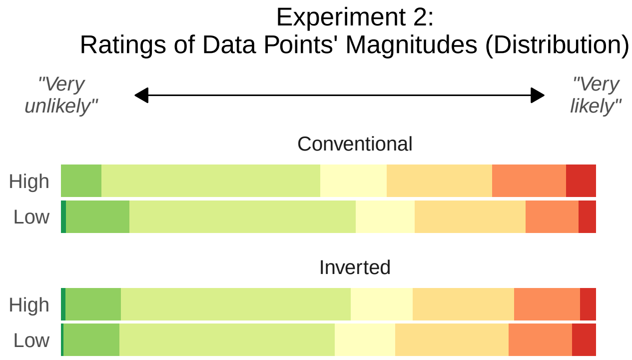

Fig. 7. Participants rated the chance of each negative event occurring on a 7-point Likert scale. The distribution of ratings, ranging from "Very unlikely" (far left, dark green) to "Very likely" (far right, red) is shown separately for each combination of the levels of each condition (axis orientation: 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.

3.4 Discussion

In E1, when using conventional charts only, we found that displaying data within different axis limits affected magnitude judgments. How-

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

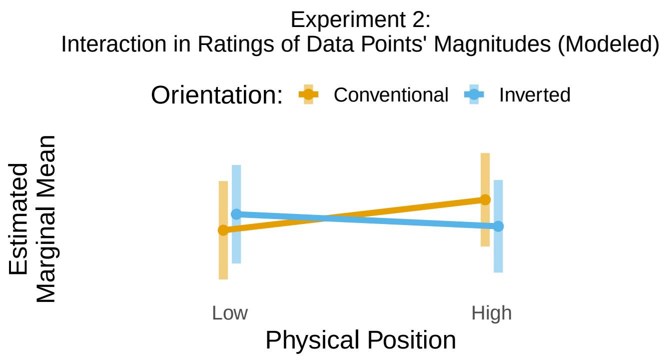

Fig. 8. Estimated marginal means for ratings of data points' magnitudes (generated by the cumulative-link mixed model). The slope for conventional charts differs from the slope of inverted charts. Thus, the effect of position on interpretation of data points' magnitudes differs according to axis orientation. Translucent bars show 95% confidence intervals.

ever, it was unclear whether judgments were based on data points' absolute positions, or their positions relative to blank space, because both would generate similar interpretations. Therefore, in E2, for half of trials, we reversed the mapping of values in physical space, so these two features would imply different magnitudes for a given value.

In E2, we replicated the primary finding from E1. In charts with conventional axis orientations, the same data points elicited different chance judgments when presented at different positions. These differences were consistent with magnitudes implied by data points' absolute positions and their positions relative to blank space. However, in charts with inverted axis orientations, the same pattern was not observed. Therefore, we can conclude that interpretations of magnitude are affected by a chart's physical arrangement of values. The pattern of differences in magnitude judgments for data points presented at distinct physical positions depends on how axes are oriented.

Figure 8 suggests that the pattern of results for inverted charts is the reverse of the pattern for conventional charts. However, our analysis indicates that the same data points did not elicit significantly different magnitude judgments when presented at different positions in *inverted* charts. Therefore, we cannot conclude from this analysis that magnitude judgments are driven solely by data points' positions relative to blank space. The lack of significant difference is likely due to a lack of experimental power. An additional experiment is required to confirm whether there is a genuine difference.

4 Experiment 3

4.1 Introduction

The interaction in E2 revealed that the influence of position on magnitude judgments depends on how different numerical values are arranged in a chart (axis orientation). The pattern of responses in inverted charts appeared to be the inverse of the pattern for conventional charts. This suggests that participants may not have based inferences about magnitude on data points' absolute positions, but on their positions relative to blank space. However, the absence of a significant difference between ratings for data points at different positions in inverted charts prohibits the conclusion that interpretations are driven entirely by comparisons with plausible values implied by blank space.

It is possible that no significant effect was detected due to insufficient experimental power. Unlike E1, with 150 participants in a single-factor design, E2 involved 120 participants in a Latin-squared 2 x 2 design. Despite an increase in the number of experimental scenarios (from 20 to 24), there were still fewer observations for each unique condition (3000 in E1 vs. 720 in E2).

In E3 we increase the experimental power and focus only on inverted charts. This will provide a clearer account of how magnitude is interpreted in inverted charts, furthering understanding of the mechanism by which axis ranges influence interpretations of magnitude.

We outlined in our pre-registration what different patterns of magnitude ratings would signal about the mechanisms used to interpret magnitude. Specifically, use of absolute position would be indicated by higher magnitude ratings for data points at *high* physical positions (mirroring the finding for conventional charts). Alternatively, use of position relative to blank space would be indicated by higher magnitude ratings for data points at *low* physical positions (the reverse of the finding for conventional charts).

4.2 Method

4.2.1 Materials

Materials were identical to E1, except for the inversion of the y-axis in all charts, including practice trials. There were 60 trials in total (40 experimental trials, 15 fillers, 5 attention check questions).

4.2.2 Procedure

The experiment was programmed in PsychoPy (version 2021.2.3 Peirce et al. [26]). As in E2, participants were asked to indicate their education level. One slide in the instructions explained to participants how charts with inverted axes function: *"In all graphs in this experiment, the arrow on the 'Chance' axis points downwards, meaning the numbers get bigger as the axis goes down."*. Otherwise, the procedure was identical to E1.

4.2.3 Design

As in E1, we employed a repeated-measures, within-participants design. Participants encountered each experimental scenario twice: once with data presented at a high physical position and once with data presented at a low physical position.

4.2.4 Participants

The experiment was not advertised on Prolific.co to those who had participated in E1 or E2, or those who signed-up to Prolific.co after 24th July 2021. Normal or corrected-to-normal vision and English fluency were required for participation.

Data were returned by 161 participants. Per pre-registered exclusion criteria, 10 participants' submissions were rejected because they answered more than two of 10 attention check questions incorrectly. One additional participant was excluded from the final dataset because they exceeded the maximum completion time (87 minutes). This left a total of 150 participants whose submissions were used for analysis: (60.00% male, 40.00% female). Mean age was 29.64 (SD = 9.56)[4]. 100% had completed at least secondary education. The mean graph literacy score was 21.87 (SD = 4.28). Participants whose submissions were approved were paid £3.45, and average completion time was 24 minutes. Ethical approval was granted by The University of Manchester's Division of Neuroscience & Experimental Psychology Ethics Committee (Ref. 2021-11115-20745).

4.3 Analysis

Figure 9 plots the distribution of participants' ratings of data points' magnitudes, for data points presented at different physical positions in inverted charts.

A likelihood ratio test reveals that a model including physical position as a fixed effect explains significantly more variability in ratings of data points' magnitudes than a model which does not include physical position as a fixed effect ($\chi^2(1) = 46.45$, $p < .001$). Data points' magnitudes were rated as greater when those data points were presented at low physical positions, compared to when the same data points were presented at high physical positions ($z = 6.80$, $p < .001$). This model employed random intercepts for each scenario. This effect remained when adjusting for participants' graph literacy scores ($z = 6.83$, $p < .001$). Estimated marginal means for these ratings are plotted in Figure 10.

For ratings of the severity of consequences, a likelihood ratio test reveals that a model including physical position as a fixed effect did not

[4]Age data was unavailable for two participants, but was available for all other participants in the dataset.

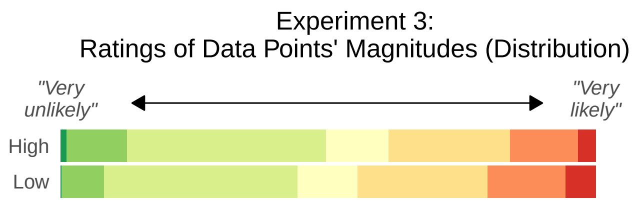

Fig. 9. Participants rated the chance of each negative event occurring on a 7-point Likert scale. The distribution of ratings, ranging from "Very unlikely" (far left, dark green) to "Very likely" (far right, red), is shown separately for charts where values were presented at a high physical position (top) and a low physical position (bottom). Note that data points at high physical positions elicited a larger proportion of ratings on the left-hand side (which represents smaller magnitudes), compared to data points at low physical positions, which elicited a larger proportion of ratings on the right-hand side (representing greater magnitudes).

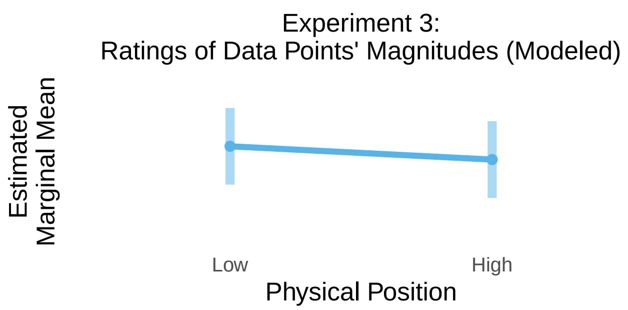

Fig. 10. Estimated marginal means for ratings of data points' magnitudes (generated by the cumulative-link mixed model). Magnitudes were rated as greater when data points in inverted charts were presented at low physical positions. Translucent bars show 95% confidence intervals.

explain significantly more variability in ratings than a model without this fixed effect ($\chi^2(1) = 3.40$, p = .065). This model employed random intercepts for each scenario, plus random intercepts and by-position slopes for each participant. This finding remained when adjusting for participants' graph literacy scores (z = 1.85, p = .064).

Models employing flexible decision thresholds (as above) were superior to models employing equidistant thresholds, for ratings of the magnitude of data points themselves ($\chi^2(4) = 752.74$), and ratings of the severity of consequences (χ^2 (4) = 177.78, $p < .001$).

4.4 Discussion

When viewing charts with inverted axes, participants judged data points' magnitudes according to whether accompanying blank space implied the existence of higher or lower plausible values. Participants ignored conventional associations between position and magnitude to interpret magnitude in the context of the chart.

In the previous experiment (E2), we did not observe a significant difference between magnitude ratings for data points at different positions in inverted charts, although the pattern was consistent with the use of blank space in the interpretation of the plotted data. However, E3, with increased experimental power, demonstrates that such a difference is statistically significant.

E2 involved switching between conventional and inverted charts, whereas E3 presented inverted charts in isolation. However, the differences in estimated marginal means for inverted charts, which represent the differences in ratings of data points' magnitudes when presented at different positions, are almost identical for these two experiments (E2: 0.34; E3: 0.33). This suggests inverted charts were not treated differently in the different experiments. Therefore, the presence or absence of switching should not prohibit the use of E3's data in explaining E2's interaction.

In light of this, we can interpret the results of E2 more easily. The same data points, presented at the same positions in a chart, convey different magnitudes depending on how they compare to plausible values implied by blank space. Viewers do not draw upon simple associations between vertical position and magnitude, but recognize the context in which values are plotted.

5 General Discussion

Over three experiments, we demonstrate how judgments of data points' magnitudes are influenced by the presence of blank space in a chart. Regardless of their physical positions, data points were associated with greater magnitudes when they were numerically greater than the plausible values represented by blank space. This was observed for charts with both conventional and inverted axes. This highlights viewers' sensitivity to context in the interpretation of information in data visualizations, suggesting designers should consider this aspect when creating charts.

When comparing data points within a single chart, it is appropriate to infer that data points which appear at different positions between two axis limits have distinct magnitudes. The results we report indicate that magnitude judgments can vary when *the same value* appears at different positions between two axis limits. Interpretation of an absolute value is biased by its relative position.

The impact of surrounding information on assessments of data is an example of a framing effect. We illustrate that this effect occurs in the absence of contrasting data points: the presence of blank space is sufficient for implying the relative status of plotted data.

The present data complement findings from prior research on y-axis truncation, which has found that the choice of axis limits can impact interpretation of data. The results we report reinforce the notion that the amount of blank space surrounding plotted values influences viewers' impressions of those values. While previous investigations have shown that y-axis limits affect *comparisons* of plotted values [7, 35, 37], the present findings show that they also affect *magnitude judgments*.

A previous study addressing a similar question also concluded that a data point's location within a range of values affects interpretation of its magnitude [29]. The present set of experiments builds upon this research by identifying the mechanism behind this effect and removing the confound of variable axes ranges. It also extends the finding beyond a single scenario to a wider range of situations, and separately analyses specific judgments, rather than using a combined measure, to verify that different presentations affect judgments of the specific variable plotted in a chart.

This set of experiments was not concerned with endorsing or opposing inverted charts; the sole function of these charts was in distinguishing competing explanations. However, when explicit instruction was provided, our data provide evidence of comprehension, contrary to the typical finding of misinterpretation resulting from associating higher positions with higher values [36, 25].

Visualization rhetoric involves presenting numerical information in a way that provokes a particular interpretation [16]. The manipulation of visualization components examined in the present set of experiments is related to two rhetorical strategies: *axis thresholding* and *contrast*. The former is an instance of 'information access' rhetoric, and involves setting an axis range that provides an incomplete picture of the data. The latter is an instance of 'mapping' rhetoric, and employs visual properties to promote comparisons.

We did not find consistent evidence that assessments of the severity of consequences are affected by the positioning of values representing the chance of events occurring. Prior research has found that probability estimates change as a function of outcome magnitude [14, 13] and that outcome magnitude estimates change as a function of event probability [19]. However, whereas prior research focused on the potency of an event, we asked participants to evaluate another feature: the severity of its consequences. How affected parties are impacted by an event is one step removed from a core component of risk, outcome magnitude. In addition, unlike prior work which substantially manipulated underlying scenarios, our more subtle manipulation retained the same probability values, changing only the surrounding context. The effect of relative position on interpretation of chance data does not consistently extend to judgments about the severity of consequences.

Adjusting for data visualization literacy did not remove the influence of axis range on interpretations. Yang et al. [37] also observed that data visualization literacy could not sufficiently explain variance in the degree of bias caused by y-axis truncation. This measure captures comprehension of the conventions of data visualization, indicating receipt of elementary instruction [23]. Therefore, it is perhaps better suited to measuring ability to decipher more complicated designs, but is not well-placed to predict susceptibility to differences in presentation format [37].

5.1 Implications for Visualization Design

This finding highlights an opportunity for data visualization designers to creatively construct axes for dramatic effect. Introducing blank space when setting axis limits allows designers to persuasively convey large or small magnitudes. However, even those avoiding creative use of blank space should be sensitive to our finding that axis ranges are likely to be considered representative of relevant values for assessing the magnitude of plotted data. Designers should consider what is *not* plotted and reflect on the impression(s) of magnitude resulting from their choice of axis limits. To avoid misleading displays, axes should present appropriate values. Like Correll et al. [7], we acknowledge that there is no objectively correct method for achieving this. Ultimately, the designer decides what context is appropriate, based on the chart's purpose and content. This may involve taking into account historical data, comparable scenarios, established baselines, current objectives, *etc.*. Our findings are also relevant for assessing the quality of data visualizations; one should consider whether a chart appropriately portrays magnitude, in addition to standard considerations.

Setting an axis range that extends far beyond the range of the plotted data impacts discrimination ability [35], and may distract attention from meaningful variance within the data. Witt recommends setting an axis range to 1.5-2 times the plotted data's standard deviation. This guidance is broadly consistent with our suggestions in its recommendation that axis limits should take into account relevant values to provide context. The present experiment has demonstrated that magnitude is communicated by the relative position of data points within the space of all plausible values.

When following Witt's [35] suggestions, data points' positions are determined solely by the size of the numerical difference between two conditions. A large difference between conditions would result in data points being located near the two extremes of the chart, which may capture genuine small and large magnitudes. At other times, applying Witt's guidance will create an inaccurate impression of individual magnitudes. For example, with a small difference between conditions, no data points will be displayed near the extremes, even though they may be genuinely large or small when considered within a larger context. This occurs because Witt's guidance was created for the sole purpose of managing bias and sensitivity when comparing two conditions (in fields with standardized effect sizes). Accordingly, setting axes which provide context for *individual* magnitudes, is not considered pertinent. Again, designers must consider their dataset and the message they intend to relate in order to reach a trade-off between suitable communication of variability and individual magnitudes. A possible compromise may involve displaying values against blank space to convey magnitude in context, and also in a focused display to facilitate comparisons between values. This resembles an approach for communicating differences discussed by Correll et al. [7], and reported to benefit users by Ritchie et al. [28]. Its suitability for conveying magnitude should be investigated in future work.

5.2 Limitations

To avoid likelihood of misinterpretation, participants were given instructions on how to read inverted charts. This may have suppressed a spontaneous interpretation of magnitude, based on physical position, in favor of a learned interpretation. Our investigation therefore only explains how viewers interpret magnitude when they know how to interpret a given chart.

In addition to associations between vertical position and magnitude, vertical position is also a common conceptual metaphor for emotional valence. Lower physical positions are typically associated with negative valence and higher physical positions with positive valence. Woodin et al. [36] found that comprehension is facilitated when the physical arrangement of data is consistent with the conceptual metaphor for valence, but that associations between vertical position and numerical magnitude affect interpretations more strongly. In the present set of experiments, charts displayed negative outcomes, so data were aligned with the conceptual metaphor for valence in inverted charts, and misaligned in conventional charts. Participants evidently did not use valence metaphors to interpret values in conventional charts; this would have produced the opposite pattern of results to those observed. The simplest explanation for our results is that participants relied on relative position when interpreting both conventional and inverted charts, rather than sometimes generating inferences based on a conceptual metaphor for valence.

In analyses employing graph literacy as a co-variate, graph literacy scores were calculated as the average of five Likert scale responses. This means that responses to graph literacy questions were modeled as continuous data, whereas Likert scale ratings from experimental trials were modeled as ordinal data. This approach was used by the scale's developers [10], but is not the most appropriate method [21].

5.3 Conclusion

The position of data points in a chart affects interpretation of how big or small their values are. We demonstrate that this relationship between physical position and inferences about magnitude critically depends on whether accompanying blank space represents higher or lower alternatives to the plotted data. Viewers take into account the context in which data appears, even when comparison values are not explicitly displayed. Axis limits and blank space warrant consideration from data visualization designers.

Acknowledgments

Duncan Bradley was supported by the Economic and Social Research Council (Grant Number ES/P000665/1). This work was supported in part by a BPS Cognitive Section Postgraduate Rapid Project Grant. We thank Jen McBride and Paul Warren for comments on an earlier draft, and Paul Stott for assistance with manuscript formatting.

References

[1] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278, 2013. doi: 10.1016/j.jml.2012.11.001.

[2] M. A. Borges and B. K. Sawyers. Common verbal quantifiers: Usage and interpretation. *Journal of Experimental Psychology*, 102(2):335–338, 1974. doi: 10.1037/h0036023.

[3] G. D. A. Brown, J. Gardner, A. J. Oswald, and J. Qian. Does Wage Rank Affect Employees' Well-being? *Industrial Relations*, 47(3):355–389, 2008. doi: 10.1111/j.1468-232X.2008.00525.x.

[4] N. Charalambides. We recently went viral on TikTok - here's what we learned, 2021. [Online]. Available: `https://blog.prolific.co/we-recently-went-viral-on-tiktok-heres-what-we-learned/`.

[5] R. H. B. Christensen. ordinal—Regression Models for Ordinal Data, 2019. [Online]. Available: `https://CRAN.R-project.org/package=ordinal`.

[6] W. S. Cleveland, P. Diaconis, and R. McGill. Variables on Scatterplots Look More Highly Correlated When the Scales Are Increased. *Science*, 216(4550):1138–1141, 1982. doi: 10.1126/science.216.4550.1138.

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

[8] L. S. Elting, C. G. Martin, S. B. Cantor, and E. B. Rubenstein. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. *BMJ*, 318(7197):1527–1531, 1999. doi: 10.1136/bmj.318.7197.1527.

[9] D. Feldman-Stewart, N. Kocovski, B. A. McConnell, M. D. Brundage, and W. J. Mackillop. Perception of Quantitative Information for Treatment

Decisions. *Medical Decision Making*, 20(2):228–238, 2000. doi: 10.1177/0272989X0002000208.

[10] R. Garcia-Retamero, E. T. Cokely, S. Ghazal, and A. Joeris. Measuring Graph Literacy without a Test: A Brief Subjective Assessment. *Medical Decision Making*, 36(7):854–867, 2016. doi: 10.1177/0272989X16655334.

[11] M. Gattis. Structure mapping in spatial reasoning. *Cognitive Development*, 17(2):1157–1183, 2002. doi: 10.1016/S0885-2014(02)00095-3.

[12] P. Grice. Logic and Conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics Vol.3: Speech Acts*, pages 41–58. Academic Press, New York, 1975.

[13] A. J. Harris, A. Corner, and U. Hahn. Estimating the probability of negative events. *Cognition*, 110(1):51–64, 2009. doi: 10.1016/j.cognition.2008.10.006.

[14] A. J. L. Harris and A. Corner. Communicating environmental risks: Clarifying the severity effect in interpretations of verbal probability expressions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6):1571–1578, 2011. doi: 10.1037/a0024195.

[15] History, Art & Archives, U.S. House of Representatives, Office of the Historian. Black-American Members by Congress, 1870–Present. [Online]. Available: `https://history.house.gov/Exhibitions-and-Publications/BAIC/Historical-Data/Black-American-Representatives-and-Senators-by-Congress/`.

[16] J. Hullman and N. Diakopoulos. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2231–2240, 2011. doi: 10.1109/TVCG.2011.255.

[17] C. Keller, M. Siegrist, and V. Visschers. Effect of Risk Ladder Format on Risk Perception in High- and Low-Numerate Individuals. *Risk Analysis*, 29(9):1255–1264, 2009. doi: 10.1111/j.1539-6924.2009.01261.x.

[18] G. Keren and K. H. Teigen. The probability—outcome correspondence principle: A dispositional view of the interpretation of probability statements. *Memory & Cognition*, 29(7):1010–1021, 2001. doi: 10.3758/BF03195763.

[19] D. Kupor and K. Laurin. Probable Cause: The Influence of Prior Probabilities on Forecasts and Perceptions of Magnitude. *Journal of Consumer Research*, 46(5):833–852, 2020. doi: 10.1093/jcr/ucz025.

[20] R. V. Lenth. emmeans: Estimated Marginal Means, aka Least-Squares Means, 2021. [Online]. Available: `https://CRAN.R-project.org/package=emmeans`.

[21] T. M. Liddell and J. K. Kruschke. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79:328–348, 2018. doi: 10.1016/j.jesp.2018.08.009.

[22] Y. Okan, R. Garcia-Retamero, M. Galesic, and E. T. Cokely. When Higher Bars Are Not Larger Quantities: On Individual Differences in the Use of Spatial Information in Graph Comprehension. *Spatial Cognition & Computation*, 12(2-3):195–218, 2012. doi: 10.1080/13875868.2012.659302.

[23] Y. Okan, M. Galesic, and R. Garcia-Retamero. How People with Low and High Graph Literacy Process Health Graphs: Evidence from Eye-tracking: Graph Literacy and Health Graph Processing. *Journal of Behavioral Decision Making*, 29(2-3):271–294, 2016. doi: 10.1002/bdm.1891.

[24] Y. Okan, E. R. Stone, J. Parillo, W. Bruine de Bruin, and A. M. Parker. Probability Size Matters: The Effect of Foreground-Only versus Foreground+Background Graphs on Risk Aversion Diminishes with Larger Probabilities. *Risk Analysis*, 40(4):771–788, 2020. doi: 10.1111/risa.13431.

[25] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini. 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*, pages 1469–1478, Seoul Republic of Korea, 2015. ACM. doi: 10.1145/2702123.2702608.

[26] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203, 2019. doi: 10.3758/s13428-018-01193-y.

[27] R Core Team. R: A Language and Environment for Statistical Computing, 2021. [Online]. Available: `https://www.R-project.org/`.

[28] J. Ritchie, D. Wigdor, and F. Chevalier. A Lie Reveals the Truth: Quasimodes for Task-Aligned Data Presentation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Glasgow Scotland Uk, 2019. ACM. doi: 10.1145/3290605.3300423.

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

[30] Y.-T. Sung and J.-S. Wu. 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, 2018. doi: 10.3758/s13428-018-1041-8.

[31] B. G. Taylor and L. K. Anderson. Misleading Graphs: Guidelines for the Accountant. *Journal of Accountancy*, 162(4):126–135, 1986.

[32] A. Tversky and D. Kahneman. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157):1124–1131, 1974. doi: 10.1126/science.185.4157.1124.

[33] B. Tversky. Cognitive Principles of Graphic Displays. In *Proceedings of the AAAI 1997 Fall Symposium on Reasoning with Diagrammatic Representations*, pages 116–124, Menlo Park, CA, 1997. AAAI Press,.

[34] C. C. Voeten. buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects, 2022. [Online]. Available: `https://CRAN.R-project.org/package=buildmer`.

[35] J. K. Witt. Graph Construction: An Empirical Investigation on Setting the Range of the Y-Axis. *Meta-Psychology*, 3, 2019. doi: 10.15626/MP.2018.895.

[36] G. Woodin, B. Winter, and L. Padilla. Conceptual Metaphor and Graphical Convention Influence the Interpretation of Line Graphs. *IEEE Transactions on Visualization and Computer Graphics*, 28(2):1209–1221, 2022. doi: 10.1109/TVCG.2021.3088343.

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

[38] B. J. Zikmund-Fisher, A. Fagerlin, and P. A. Ubel. What's Time Got to Do with It? Inattention to Duration in Interpretation of Survival Graphs. *Risk Analysis*, 25(3):589–595, 2005. doi: 10.1111/j.1539-6924.2005.00626.x.