

Top of the Charts: The Influence of Position and Axis Range on Interpretations of Magnitude

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

Abstract—Type abstract here

1 INTRODUCTION

type intro here

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%, but may be considered small for the chance that a laboratory test can successfully detect a disease.

In charts, numerical axes often provide some contextual cues for judging the magnitude of plotted values (how large or small they are). The range of values on an axis provides a frame of reference for assessing whether a data point is large or small. A bar chart produced by the New York Times, which plots over time the number of black members of the U.S. senate, provides a striking illustration (Figure 1). Unusually, the continuous y-axis does not terminate just above the highest 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 user'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, assessments of magnitude may be influenced by the appearance of plotted values only, and not through contrast with blank space. Users may simply judge data points at higher positions as 'high' and those lower down as 'low', regardless of any other cues which might communicate magnitude. 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 Judgements

Empirical evidence demonstrates that judgement of a value's magnitude depends on its relationship to a grand total or to surrounding values. This affects 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 & 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, Gardner, Oswald, & Qian, 2008).

- Author one is with Uni of X. Email: author1@uniofX.com
- Author one is with Uni of X. Email: author1@uniofX.com
- Author three is with Uni of X. Email: author1@uniofX.com

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.xxx/TVCG.201x.xxxxxxx

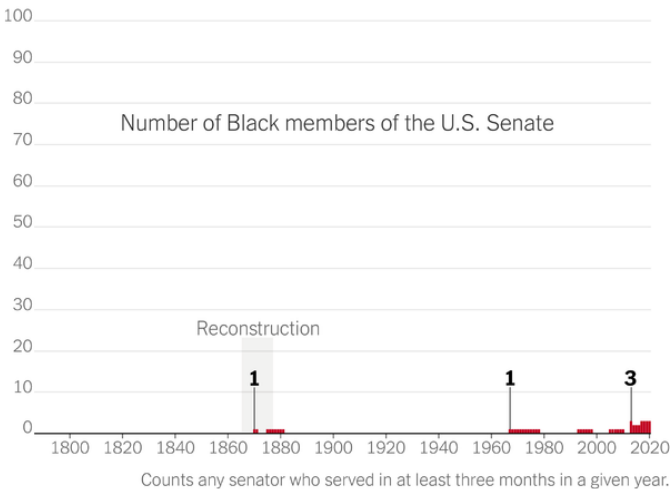


Fig. 1. In this chart, the y-axis limit is the largest possible value, rather than the largest observed value, so plotted values appear to have particularly small magnitudes.

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 magnitude of the 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 (Cleveland, Diaconis, & McGill, 1982). 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 (Pandey, Rall, Satterthwaite, Nov, & Bertini, 2015). Correll et al.'s (2020) experiments found that greater truncation resulted in higher effect-size judgements 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. (2020) found no reduction in effect size judgements when truncation was communicated using graphical techniques (e.g., axis breaks and gradients). The unavoidable consequence, they suggest, is that designers' choices will influence users' interpretations whether axes are truncated or not.

Choosing an appropriate axis range involves a trade-off between bias (over-reliance on visual appearance of differences) and sensitivity (capacity to visually recognise differences that exist). 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 (Witt, 2019). 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 (0-100). This provides further evidence of a powerful association between the appearance of

data, when plotted, and subjective interpretations of these data.

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 very small sample sizes (Witt, 2019), thus resulting in a lack of experimental power. Low powered studies tend to overestimate magnitude of true effect sizes. Participants' ratings of the difference between two bars consistently provided evidence of the exaggerating effects of y-axis truncation. Yang et al. (2021) discussed various potential mechanisms behind truncation effects and noted that the effect appears to be largely automatic, so may function like an anchoring effect, where a numerical judgement is influenced by a reference point (Tversky & Kahneman, 1981). Another explanation 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. Therefore, a user will assume that a difference must be genuinely large if it appears large, else it would not be presented that way. Effective visualisations should be designed so a user's instinctive characterisation of the data corresponds closely to their interpretation following a more detailed inspection (Yang et al., 2021).

1.3 Effects of Axis Limits on Extraction of Values

The above research consistently demonstrates that relationships between data points are 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 index of magnitude. For example, children appear to intuitively 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 (Tversky, 1997).

Research on data visualisations demonstrates the strength of the association between magnitude and vertical position: inversions of this mapping in charts can lead to misinterpretations (Okan, Garcia-Retamero, Galesic, & Cokely, 2012; Pandey et al., 2015; Woodin et al., 2021). Furthermore, other research has found that 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 (Taylor & Anderson, 1986). Sandman et al. (1994) investigated assessments of magnitude in risk ladders, where greater risks are presented at physically higher positions on a vertical scale. To assist evaluation of risk magnitudes for asbestos exposure (a relatively unfamiliar hazard), a scale for smoking frequency (a more familiar hazard) was presented alongside for reference. Participants rated the threat of asbestos 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 for this is far from strong. Taylor and Anderson (1986) did not disclose how judgements were elicited, or provide details of their sample size. Sandman et al. (1994) only explored responses towards one specific risk (asbestos), and each participant only took part in a single trial. In addition, the 'threat' measure they used was a composite of several separate ratings, which prevents 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 data elicit different inferences about magnitude, and to provide recommendations for best practice, stronger evidence is required, as is an investigation into the possible cognitive mechanisms involved in the generation of these inferences.

1.4 The Present Study

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

All data visualisations used in the present set of experiments displayed the probability of the occurrence of negative events. This provides participants with a purpose; 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 risk communication, and manipulating aspects of a chart can change interpretations of the risks displayed (Elting, Martin, Cantor, & Rubenstein, 1999; Feldman-Stewart, Kocovski, McConnell, Brundage, & Mackillop, 2000; Keller, Siegrist, & Visschers, 2009; Okan, Stone, Parillo, Bruine de Bruin, & Parker, 2020; Zikmund-Fisher, Fagerlin, & Ubel, 2005).

Risk events are composed of two core components: 1) probability of occurrence and 2) severity of harm. Individuals' assessments of probability and severity are not necessarily independent. A particular event is perceived as more likely when it is described as having more severe outcomes (Harris & Corner, 2011; Harris et al., 2009). In a similar manner, a particular event is associated with more substantial consequences when it is described as more probable (Kupor & 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 & Teigen, 2001). According to this account, probabilities can occasionally provide meaningful indications of 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 outcome probability, assessments of outcome severity may also differ between conditions. Collecting separate judgements of outcome probability and outcome severity for each scenario provides a clearer picture how the manipulation affects each aspect of participants' representations of risk. Use of Likert scales (with discrete options) rather than visual analogue scales (with continuous options; Sung & Wu, 2018) prevents participants from simply mapping probability percentages directly onto a linear scale. We also administered a data visualisation literacy measure, to determine the degree to which our manipulation affects interpretations after we have accounted for differences in graph literacy. Responses to visualisations which violate graphical conventions by using atypical scales suggest that individuals with lower graph literacy are more likely to draw on data points' physical positions when making inferences about their magnitudes (Okan, Galesic, & Garcia-Retamero, 2016; Okan et al., 2012).

2 EXPERIMENT 1

2.1 Introduction

Our initial experiment sets out to investigate whether changing axis limits affects the interpretation of data points' magnitudes. For the different version 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 probability magnitude ratings and/or severity magnitude ratings would be greater when data points were presented at higher positions, compared to when the same data points were presented at lower positions. Pre-registration of this hypothesis is available at <https://osf.io/23wpm>

2.2 Methods

2.2.1 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 (Prolific, 2021), however, data for this experiment were collected in May 2021.

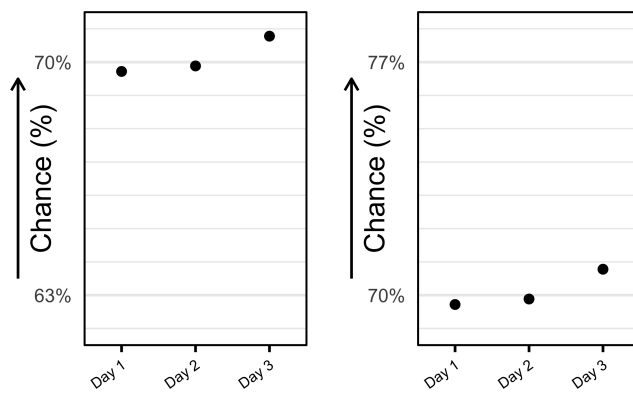


Fig. 2. Example Charts. The 'high' condition (left) presents data points near the top of the chart; the 'low' condition (right) presents the same data points near the bottom of the chart.

Normal or corrected-to-normal vision and English fluency were required for participation. Data were returned by 160 participants. Per pre-registered rejection criteria¹, 10 participants' submissions were rejected because they answered more than two of 10 attention check questions incorrectly². 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 ($SD = 12.47$)³. Participants' graph literacy scores consisted of the average of responses on five six-point scales; the mean graph literacy score was 4.26 ($SD = 0.92$). Participants whose submissions were approved were paid £3.55. Ethical approval was granted by The University of Manchester's Division of Neuroscience & Experimental Psychology Ethics Committee (Project Number:).

2.2.2 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.

Each dot plot displayed the probability (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' condition: Figure 2, left) or in the bottom third of the chart ('low' condition: Figure 2, right). The dataset differed for each distinct scenario, but was identical for the two conditions associated with a given scenario. In filler trials ($n = 15$) and attention check trials ($n = 5$), all three 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 edge of the chart was labelled with a percentage value, as was the gridline 1.5 percentage points below the top edge of the chart.

2.2.3 Procedure

The experiment was programmed in PsychoPy (Peirce, 2019) and hosted on pavlov.org. Participants were instructed to complete the

¹<https://osf.io/23wpm>

²Attention check questions are described in the Procedure.

³Age data failed to save for one participant, but was available for all other 149 participants in the dataset.

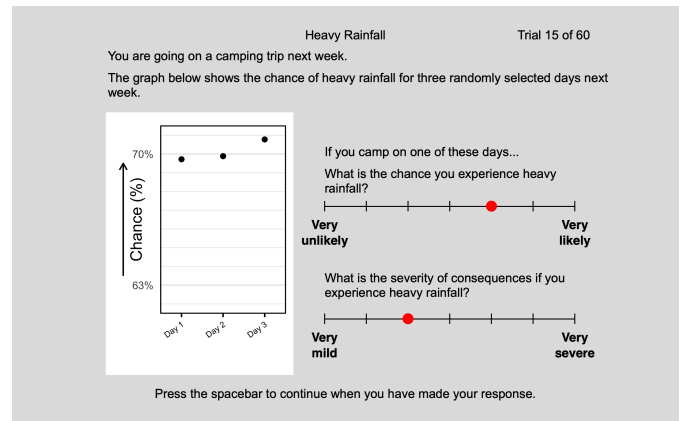


Fig. 3. Example Trial

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 graph literacy scale (Garcia-Retamero et al., 2016). They were reminded that the experiment involved information about risks, and could cause distress, so they were entitled to withdraw from the experiment at any time. Following this, instructions explained that their task involved assessing the probability 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: one rating for the probability of the negative event occurring; and one rating of the severity of the consequences if the 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 used to collect ratings had two anchors at the extremes, but all other options were unlabelled. The leftmost option in the 'probability' 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 (see 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 as follows:

What is the chance response specified above?

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.

2.2.4 Design

The experiment used a within-participants design, with each participant responding to both versions of each chart. The independent variable, vertical position, had two levels: *high* (data points presented in the upper third) and *low* (data points presented in the lower third). In each trial, responses for the two dependent variables were elicited:

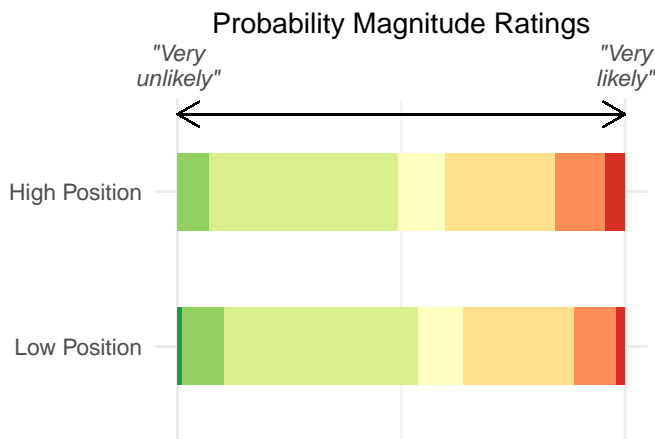


Fig. 4. Participants rated the probability of each negative event occurring on a 7-point Likert scale. The distribution of ratings, ranging from "Very unlikely" (far left, dark green) to "Very likely" (far right, red), is shown separately for charts where values were presented at a High Position (top) and a Low Position (bottom). Note that a greater proportion of values in the High Position condition were rated as representing higher probabilities (right-hand side) than values in the Low Position condition, where a greater proportion of values were rated as representing lower probabilities (left-hand side).

participants rated the probability and severity of the negative event on seven-point Likert scales.

Materials were divided into two lists. One list contained five experimental items in the 'high' condition, five experimental items in the 'low' condition, eight filler items, and two attention check questions. The other list contained the alternate versions of each of the experimental items, seven different filler items, and three different attention check questions. The order of the two lists was counterbalanced across participants, and within each list, items were presented in a random order.

2.3 Analysis

Analyses were conducted using R (version 4.1.2, R Core Team, 2021). 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 another, but do not record the magnitude of this difference. Therefore, Likert scales do not capture values from latent distributions (e.g., mental representation of probability) 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 can be inappropriate to analyse Likert scale data with metric models, such as linear regression (Liddell & Kruschke, 2018). Throughout this paper, we construct cumulative link mixed-effects models, using the *ordinal* package (version 2019.12-10, Christensen, 2019) to analyse magnitude ratings.

Selection of random effects structures for experimental mixed-effects models was automated using the *buildmer* package (version 2.3, Voeten, 2022). The maximal model structure included by-participant and by-item random intercepts, as well as by-participant and by-item slopes for the fixed effect (Barr et al., 2012). From this formula, *buildmer* first identifies the most complex model which can successfully converge, prioritising the terms which explained the most variance in the data, then eliminates terms which do not provide significant contributions (assessed using likelihood ratio tests).

Figure 4 shows the distribution of probability magnitude ratings for data points presented at high and low positions, across all experimental items.

A likelihood ratio test reveals a significant difference between the

distribution of probability magnitude ratings in the 'high position' and 'low position' conditions: ($\chi^2(1) = 74.21, p < .001$). Data points at high positions elicited higher ratings than the same data points at low positions.

There was also a significant difference between severity magnitude ratings in the 'high position' and 'low position' conditions: ($\chi^2(1) = 6.16, p = .013$). Severity magnitude ratings were higher when data points representing probabilities were presented at higher physical positions, compared to when they were presented at lower physical positions.

We also ran two additional analyses, to test whether or not these results can 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 probability magnitude ratings ($z = -8.57, p < .001$) or severity magnitude ratings ($z = -2.51, p = .012$).

2.4 Discussion

Participants rated the magnitudes of data points when those data points were presented near the top of the chart, compared to when the same data points were presented near the bottom. Modelling differences in participants' graph literacy in addition to our experimental manipulation did not remove the influence of our manipulation on interpretations.

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.

Both the outcome probability and outcome severity ratings for the data plotted on the charts were affected by the manipulation of axis limits and data points' positions, even though the charts only displayed probability information. This mirrors 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 (Kupor & Laurin, 2020).

It is unclear whether the effects of different axis ranges on interpretations of magnitude are driven by an association between *absolute* position and magnitude, or an association between *relative* position and 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). 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 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 are positioned near the *top* of a graph and substantial blank space appears *below* them, compared to when the same data points are positioned near the *bottom* of a graph, with substantial blank space *above*.

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

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 recognised the potential for larger values than those observed, consequently associating plotted data points with smaller magnitudes.

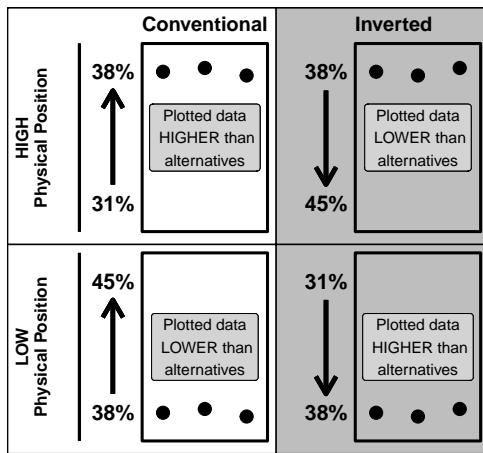


Fig. 5. Rationale for E2: Distinguishing the Roles of Absolute and Relative Position. In charts with conventional axis orientations (left column), there is congruity between data points' absolute positions and their relative positions in the chart. In charts with inverted axis orientations (right column), 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 alternatives. But at the same absolute positions, in inverted charts, the same values are relatively lower than alternatives (top right).

E1 does not provide means of differentiating these competing explanations. Drawing inferences from data points' absolute positions would orient magnitude judgements in the same direction as relying on 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: increasing *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. Therefore, inferences invoking blank space would generate the opposite impressions to inferences invoking data points' physical positions only. This is illustrated in Figure 5. In E2, we manipulate data points' physical positions by changing axis limits (as in E1), and also manipulate axis orientation, by employing conventional and inverted axes (thus giving a 2x2 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 guided by relative positions.

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 et al., 2021). In the next two experiments, we provide explicit instruction to make participants aware that inverted charts are presented.

Pre-registration of hypotheses is available at <https://osf.io/zhe7q>. For charts with conventional axis orientations, we predicted that data points presented at higher physical positions would be associated with higher probability and severity ratings, compared to data points presented at lower physical positions (replicating results from E1). For charts with inverted axis orientations, we outlined what different patterns of magnitude ratings would signal about the mechanisms used to interpret magnitude. Use of absolute position would be indicated by higher magnitude ratings for data points at higher physical positions, whereas use of relative position would be indicated by higher magnitude ratings for data points at lower physical positions. Our main objective, expressed in the 'Analyses' section of the pre-registration, was to determine whether the difference in responses to data points at distinct positions differed depending on whether those positions were

associated with greater or smaller numbers in the context of the chart.

3.2 Method

3.2.1 Design

We employed a Latin-squared within-participants design: each combination of position and axis orientation conditions was presented to each participant, but unlike E1, each participant only viewed one version of each scenario.

3.2.2 Materials

For this experiment, we increased the number of scenarios in the experimental condition. This provides some compensation for the reduced experimental power caused by a reduction in participant numbers, and a shift to a Latin-squared design where participants only viewed one graph for each scenario (thus reducing the number of observations per participant).

Two scenarios which were fillers in E1 were used as experimental items and three extra scenarios were added for use as experimental items. One filler item was removed due to a concern about its quality (it concerned the risk to others as well as the risk to oneself). E2 used a total of 24 experiment items, 12 filler items and 5 attention check questions (41 trials in total).

3.2.3 Participants

The experiment was not made available on Prolific.co to those who had participated in E1, or those who signed-up to prolific.co after 24th July 2021 (thus prior 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. 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 ($SD = 10.45$). 100.00% had completed at least secondary education. The mean graph literacy score was 4.34 ($SD = 0.94$). Participants whose submissions were approved were paid £2.37. Ethical approval was granted by The University of Manchester's Division of Neuroscience & Experimental Psychology Ethics Committee (Project Number:).

3.2.4 Procedure

An additional slide in the instructions encouraged participants to pay attention to axis orientation:

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.

Participants also specified the highest level of education they had received, in addition to demographic questions on age and gender. Otherwise, the procedure was the same as E1.

3.3 Analysis

Figure 6 plots participants' ratings of probability, for data points presented at high and low positions, in charts with conventional axis orientations and inverted axis orientations.

For probability magnitude ratings, there was a significant interaction between position in physical space, and y-axis orientation ($\chi^2(1) = 8.22$, $p = .004$). This interaction is plotted in Figure 3.3.1.

Pairwise comparisons (with Sidak adjustment) reveal that the main effect of position in charts with conventional y-axis orientations (E1) is replicated ($z = 3.56$, $p = .001$). Data points at higher positions were

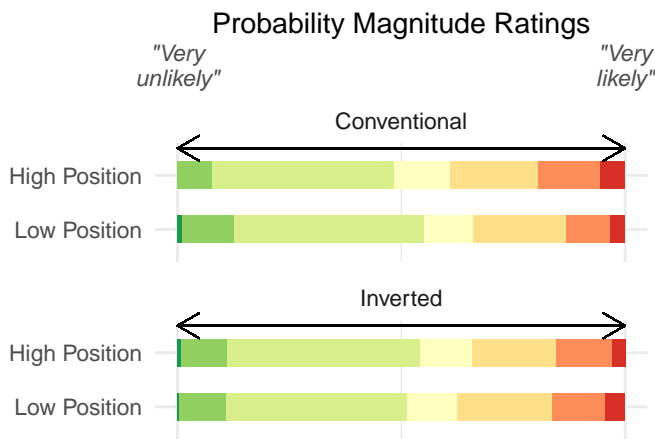


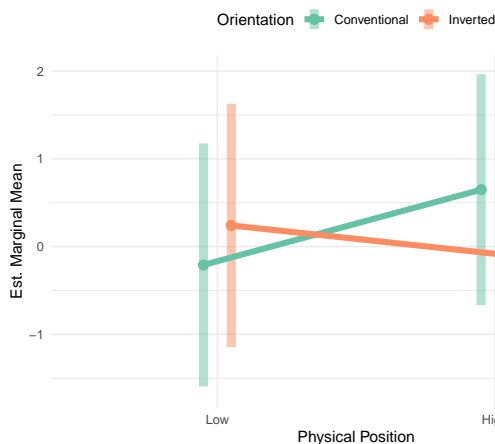
Fig. 6. Participants rated the probability of each negative event occurring on a 7-point Likert scale. The distribution of ratings is shown separately for each combination of the levels of each condition (axis orientation: conventional, inverted; data points' position: high, low). Note that the pattern 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, higher probability ratings were more common in the High Position condition, whereas when charts used inverted axes, higher probability ratings were more common in the Low Position condition.

associated with higher probability ratings than data points at lower positions.

There was no significant difference between ratings for data points plotted at different positions when inverted axes were used ($p = .512$). Therefore, we observe a different pattern of results when an inverted axis is used, compared to when a conventional axis is used.

Data points plotted at high physical positions were associated with higher ratings when a conventional axis was used than when an inverted axis was used ($-3.08, p = .008$). However, data points plotted at low physical positions were not associated with different ratings depending on the axis orientation ($2.15, p = .119$).

3.3.1 intplot



\begin{figure}
\caption{Estimated marginal means (generated by our statistical model) for responses to data points at high and low physical positions, in charts with conventional and inverted axes. The slope for conventional charts differs from the slope of inverted charts. Thus, the effect of position on interpretation of magnitudes differs according to axis orientation. Translucent bars show 95% confidence intervals.}
\end{figure}

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 also remained when controlling for participants' graph literacy scores: $z = -2.91, p = .004$.

There was also an interaction between physical position and axis orientation for severity magnitude ratings: $(\chi^2)(1) = 5.13, p = .024$. However, the main effect in severity ratings from E1, different responses to data points at different positions in conventional graphs, 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 positions in inverted and conventional charts ($-2.52, p = .047$). Modelling participants' graph literacy scores as an additional fixed effect did not eliminate this interaction: $z = -2.29, p = .022$.

3.4 Discussion

When using conventional graphs only (E1), we found that displaying data within different axis limits affected magnitude judgements. However, it was unclear whether judgements were based on data points' absolute or relative positions, because similar interpretations would result from both features. 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 magnitude judgements when presented at different positions. These judgements corresponded to magnitudes implied by data points' absolute and relative positions. 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 judgements for data points presented at distinct physical positions depends on how axes are oriented. This interaction was robust to inclusion of graph literacy as an additional effect.

There was not sufficient evidence to conclude that the same data points elicited different magnitude judgements when presented at different positions in *inverted* charts. Therefore, we cannot conclude from these data that magnitude judgements are solely driven by data points' positions relative to other values. However, Figure 3.3.1 suggests that the pattern of results for inverted charts is the reverse of the pattern for conventional charts. An additional experiment is required to confirm whether this is the case.

4 EXPERIMENT 3

4.1 Introduction

The significant interaction in E2 revealed that the influence of position on magnitude judgements depends on how a chart arranges different numerical values (axis orientation). The pattern of responses in inverted charts appeared to be the inverse of the pattern for conventional charts. This suggests that participants did not generate inferences about magnitude based on conceptual metaphors, and instead paid attention to context. However, the absence of a significant difference between high and low data points in inverted charts prohibits the conclusion that interpretations are driven entirely by comparisons with other parts of a chart.

It is possible that no significant effect was detected due to insufficient experimental power. Unlike E1, with 150 participants in repeated-measures single-factor design, E2 involved 120 participants in a Latin-squared 2x2 design. Despite an increase in the number of experimental items (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 the true nature of the processing of data in inverted charts, furthering understanding of the

⁴This represents a divergence from pre-registered analysis protocol. We tested whether controlling for participants' graph literacy scores would change whether effect(s) resulted from the manipulation, rather than testing for an interaction between the effect(s) and graph literacy scores literacy.

⁵See Footnote 3

mechanism which causes interpretations of magnitude to be affected by how data is displayed.

In our pre-registered hypotheses (<https://osf.io/t4snu>), we outlined what different patterns of magnitude ratings would signal about the mechanisms used to interpret magnitude. Use of absolute position would be indicated by higher magnitude ratings for data points at higher physical positions, whereas use of relative position would be indicated by higher magnitude ratings for data points at lower physical positions.

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 items. There were 60 trials in total (40 experimental items, 15 fillers, 5 attention check questions).

4.2.2 Participants

The experiment was not made available on Prolific.co to those who had participated in E1 or E2, or those who signed-up to prolific.co after 24th July 2021 (thus prior to the shift in participant demographics). Normal or corrected-to-normal vision and English fluency were required for participation.

Data were returned by 161 participants. Following pre-registered rejection 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 accepted and used for analysis: (60.00% male, 40.00% female). Mean age was 29.64 ($SD = 9.56$)⁶. 100.00% had completed at least secondary education. The mean graph literacy score was 4.37 ($SD = 0.86$). Participants whose submissions were approved were paid £3.45. Ethical approval was granted by The University of Manchester's Division of Neuroscience & Experimental Psychology Ethics Committee (Project Number:).

4.2.3 Procedure

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."* As in E2, participants were asked to indicate their education level. Otherwise, the procedure was as described above.

4.2.4 Design

Like E1, this experiment used a within-participants, fully repeated-measures design, with each participant responding to both versions of each chart.

4.3 Analysis

Figure 7 plots participants' probability magnitude ratings for data points presented at different positions in inverted charts.

A likelihood ratio test reveals a significant difference between probability magnitude ratings for data points at different positions in inverted charts: ($\chi^2(1) = 46.45$, $p < .001$). Ratings were higher for data points presented at lower positions. This effect remained when adjusting for participants' graph literacy scores ($z = -6.83$, $p < .001$).⁷

There was no difference between severity magnitude ratings in inverted charts: $\chi^2(1) = 3.40$, $p = .065$. This outcome was replicated when adjusting for participants' graph literacy scores ($z = -1.85$, $p = .064$).⁸

⁶Age data failed to save for two participants, but was available for all other 148 participants in the dataset.

⁷See Footnote 3

⁸See Footnote 3

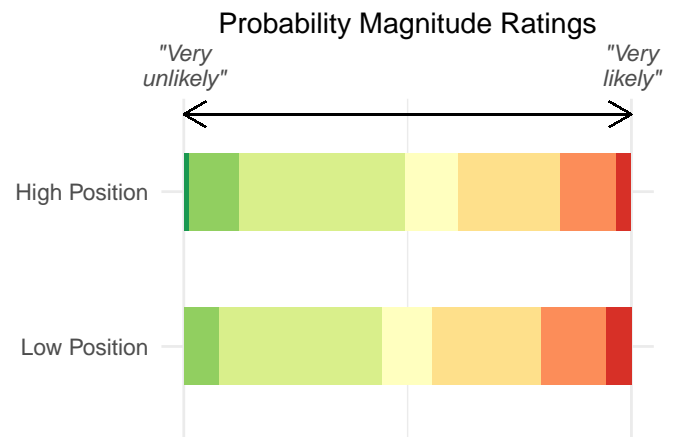


Fig. 7. Participants rated the probability of each negative event occurring on a 7-point Likert scale. The distribution of ratings, ranging from "Very unlikely" (far left, dark green) to "Very likely" (far right, red), is shown separately for charts where values were presented at a High Position (top) and a Low Position (bottom). Note that values in the Low Position condition were more frequently rated as representing higher probabilities (right-hand side) than values in the High Position condition, which were more frequently rated as representing lower probabilities (left-hand side).

4.4 Discussion

When viewing charts with inverted axes, participants judged data points' magnitude according to whether accompanying blank space implied the existence of higher or lower plausible values. They 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, for inverted graphs. We suggested that a false negative may have occurred in E2, due to insufficient experimental power to detect this difference. 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, estimated marginal means, which capture the difference in ratings for values at different positions, are almost identical across these two experiments (E2: 0.34; E3: 0.33), suggesting these 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 the interaction in E2.

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

5 GENERAL DISCUSSION

Over three experiments, we demonstrate how judgements of the magnitude of data points 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 users' sensitivity to context in interpretation of information in data visualisations, 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 judgements 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 context 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 reported that the choice of axis limits can impact interpretation of data. These results we report reinforce the notion that the amount of blank space surrounding plotted values influences users' impressions of those values. While previous investigations have shown that y-axis values affects the *comparison* of distinct values (e.g., Correll et al., 2020; Pandey et al., 2015; Witt, 2019; Yang et al., 2021), the present findings show that they also affect *judgements* of values themselves.

While previous investigations have shown that y-axis limits affect judgements of the *difference between* distinct values (e.g., Correll et al., 2020; Pandey et al., 2015; Witt, 2019; Yang et al., 2021), the present findings show that they also affect judgements of the *magnitude* of individual values.

While previous investigations have shown that y-axis limits affect the *comparison* of values (e.g., Correll et al., 2020; Pandey et al., 2015; Witt, 2019; Yang et al., 2021), the present findings show that they also affect *magnitude judgements*.

While previous investigations have shown that y-axis limits affect *comparison judgements* (e.g., Correll et al., 2020; Pandey et al., 2015; Witt, 2019; Yang et al., 2021), the present findings show that they also affect *magnitude judgements*.

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 (Sandman et al. 1994). 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 judgements, rather than using a combined measure, to verify that different presentations affect judgements 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 (Woodin et al., 2021; Pandey et al., 2015).

Visualisation rhetoric involves presenting numerical information in a way that provokes a particular interpretation (Hullman & Diakopoulos, 2011). The manipulation of visualisation 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 outcome severity are affected by the positioning of values representing event probability. This contrasts with prior research on the interplay between appraisals of probability and event magnitude which reports that probability estimates change as a function of magnitude (Harris & Corner, 2011; Harris et al., 2009) and that magnitude estimates change as a function of event probability (Kupor & Laurin, 2020). Unlike this prior work, which substantially manipulated underlying scenarios, our more subtle manipulation retained the same probability values, changing only surrounding context. The effect of relative position on interpretation of numerical information does not consistently extend to magnitude judgements about related but distinct concepts.

Adjusting for data visualisation literacy did remove the effect of the influence of axis range on interpretation. Yang et al. (2021) also observed that data visualisation 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 visualisation, indicating receipt of elementary instruction (Okan et al., 2016). There-

fore, 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 (Yang et al., 2021).

5.1 Implications for Visualisation Design

This finding highlights an opportunity for data visualisation designers to creatively construct axes for dramatic effect. Introducing blank space by setting particular 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 the relevant values for assessing the magnitude of plotted data. Designers should 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. (2020), 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 account of historical data, comparable scenarios, established baselines, current objectives, etc.. The results we report above are also relevant for assessing the quality of data visualisations; 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 (Witt, 2019), and may distract attention from meaningful variance within the data. Witt recommends setting an axis range to 1.5-2 times the standard deviation of the dataset. 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.

Following Witt's (2019) 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 standardised 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 suggested by Correll et al. (2020), and reported to benefit users by Ritchie et al. (2019). 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 favour of a learned interpretation. Our investigation therefore only explains how users interpret magnitude when they know how to interpret the chart itself.

In analyses employing graph literacy as a covariate, graph literacy scores were calculated as the average of five Likert scale responses. This means that responses to graph literacy questions were modelled as continuous data, whereas Likert scale responses to the probability and severity questions were modelled as ordinal data. This approach was used by the scale's developers (Garcia-Retamero et al., 2016), but is unlikely to be the most appropriate method.

In addition to associations between vertical position and magnitude, a common conceptual metaphor for emotional valence also uses vertical position. Lower physical positions are typically associated with negative valence and higher physical positions with positive valence. Woodin et al. (2021) 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 magnitude affect interpretations more strongly. In the present set of experiments, charts displayed negative outcomes, so data was 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 - that would have produced the opposite pattern of results to those observed. So, the simplest explanation for our data is that participants relied on relative position when interpreting both conventional and inverted charts, rather than some of the time generating inferences based on a conceptual metaphor for valence.

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. Users 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 visualisation practitioners.

6 ACKNOWLEDGEMENTS

DB was supported by the Economic and Social Research Council [Grant Number ES/P000665/1]. This work was supported by a BPS Cognitive Section Postgraduate Rapid Project Grant.