Adapting CCDF Plots for Visualizing Ordinal Regression Results

Abhraneel Sarma*
Graz University of Technology

ABSTRACT

Cumulative-link ordinal regression models are an alternative approach for analysing ordinal data such as Likert items, which are widely used in Visualization (and other related fields like HCI, psychology etc.). There are many researchers who are strong proponents of this approach, as it makes less stringent assumptions about the data, compared to the more commonly used linear model or ANOVA. Yet, ordinal regression models have seen limited adoption. I posit that one possible reason for this might be due to the difficulty in visually representing the results from such models, and in communicating the key takeaways in an intuitive manner. I propose the use of (modified) Complementary Cumulative Distribution Function (mCCDF) plots to visualize the results of ordinal regression models, and demonstrate how the same takeaways that researchers present from analyses which treat ordinal data as metric can be easily communicated using mCCDFs.

Index Terms: Complementary Cumulative Distribution Functions, Likert scale, Ordinal Regression Model

1 INTRODUCTION

An ordinal variable refers to variables where the categories have a natural order, but are not metric [29] (i.e., the differences between successive levels may not be treated as equal intervals). The use of ordinal measures (e.g., surveys which elicit participants' responses on a Likert scale) are commonly used by researchers in visualization, as well as in other fields such as HCI or psychology. These measures provide a valuable way of measuring metrics such as trust in a visualization, confidence in using a new interface, preference or attitudes etc., which we wouldn't be able to capture otherwise.

Despite their widespread use, researchers tend to use a wide variety of statistical approaches for analysing Likert data, some of which makes strong and arguably unfounded assumptions about the data. For instance, the ANOVA test, frequently used for analysing Likert responses, assumes that the data is metric [13]. Others use non-parametric approaches which, while not making assumptions on the data distribution, reduces the data to relative ordered ranks [31] which may also be problematic. Instead, Syiem and Velloso [31], echoing prior calls from other fields [13], have argued for the adoption of ordinal regression models instead, as these models make less stringent assumptions regarding the data, and more directly model the data generating process.

Like Syiem and Velloso [31], I also believe that ordinal cumulative-link models are the more appropriate approach for analysing Likert data—ordinal models make much more reasonable assumptions regarding the data-generating process and generally tends to fit the data better.[1] However, I posit that the a key reason for the lack of adoption of ordinal models is the fact that **we currently do not have good approaches for visualising the results of ordinal models** which are effective at communicating the takeaways that analysts wish to communicate regarding their analysis. As prior work has used various approaches for communicating the results of Likert scale responses, I first conduct a preliminary review of reporting strategies used for ordinal data, based on a snowball sample of prior work; I summarise the key message that researchers are trying to communicate and identify key differences between when the data is treated as metric or ordinal.[2]

I find that an important benefit of a metric-based analysis of ordinal responses is the ability to report results and describe differences between conditions on the response scale (e.g., participants in condition [X], on average, reported 2 points higher trust on a 7-point Likert scale, compared to participants in condition [Y]); these differences can also be very intuitively visualized (see section 3). Analogous statements are very difficult (although not impossible) to make if the data were analyzed using an ordinal model instead. Based on these findings, I propose the use of modified Complementary Cumulative Distribution Function (mCCDF) plots to visualize the results of ordinal regression models. Even though mCCDFs are less straightforward for interpreting the results (compared to the interval plots that researchers use when treating the results as metric), they do not violate the ordinal assumption by treating the data as metric. More importantly, I demonstrate how mCCDFs allow an analyst to communicate the analogous information using ordinal models that metric-based analyses of ordinal data affords, which can be further reinforced through annotations (section 4).

2 PRELIMINARIES

To make the discussion throughout this paper more concrete, I implement and use the results of an ordinal regression model. I introduce the dataset and the the model below (see also Appendix B)

2.1 The Dataset

To demonstrate the mCCDF visualization, we first need an ordinal model. I use a dataset on moral intuition [6, 16] which contains the results of an experiment where participants are presented with different scenarios involving the *trolley problem*. The scenarios manipulate three principles—*action*, *intention* and *contact*. In this experiment, participants are presented with a number of scenarios where one or more of these principles apply; thus whether a principle applies in a particular scenario is the experimental variable. Participants' responses are recorded on a 7-point scale where higher values indicate greater moral permissibility. For a more detailed discussion on the experiment, I would point readers to [6, 16].

2.2 The Ordinal Regression Model

I fit a Bayesian[3] cumulative-link model using the probit link function on the *moral intuition* dataset using `action`, `intention` and `contact` as predictors. The cumulative model assumes that an observed ordinal variable, Y, originates from a categorization of a

*e-mail: asarma@tugraz.at

[1] However, how often ordinal models lead to different conclusions in practice is unclear, and I elaborate more on this in Section 4.4

[2] I do not look at non-parametric responses because I think that approach represents the worst of both worlds—i.e., makes different unfounded assumptions (i.e., by converting the data into ranks, it flattens the difference between a ordered dataset of {5, 2, 3} and {7, 2, 3} as both get ranked as {3, 1, 2} [31], and remains confusing to interpret.

[3] The complete analysis can be found in `supplement` ▶ RScript ▶ `01-bayesian.Rmd`. I also show how to create the corresponding visualizations for a frequentist analysis in `supplement` ▶ RScript ▶ `02-frequentist.Rmd`

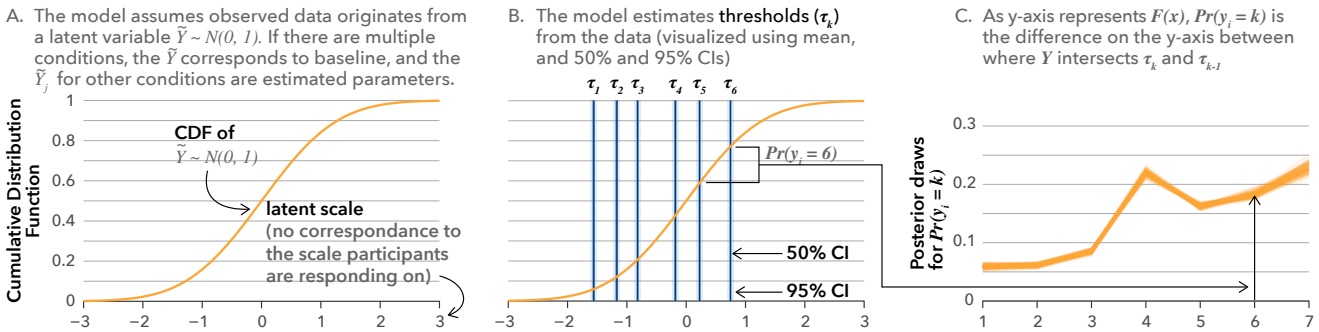

Figure 1: Illustration of how a cumulative probit-link model estimates thresholds of a latent variable (B), which can then be used to estimate (posterior) probabilities for the prevalance of each rating (C).

latent continuous variable, $\tilde{Y}$ [3, 31], which follows the unit normal distribution (Figure 1A). To achieve this categorization, the model then assumes that there are k thresholds τ_k which categorize the latent variable $\tilde{Y}$ into $k+1$ *ordered* categories (Figure 1B). The thresholds τ_k, are on the scale of the latent variable, and are estimated from the data:

$$\Phi^{-1}(\mathbb{P}(y_i \leq k)) = \tau_k$$

where Φ is the probit function[4] corresponding to the unit normal distribution. We can estimate the probability of Y being equal to a category k as $\mathbb{P}(Y = k) = \Phi^{-1}(\tau_k) - \Phi^{-1}(\tau_{k-1})$. This is extended to a regression model by assuming that the different conditions will have different underlying latent distributions $\tilde{Y}_j$ (see Figure 1).

3 Visualizing Ordinal Data

To get a better understanding of what researchers want to communicate from their analysis of ordinal data, I conducted an informal survey of prior work in visualization. I began with a few initial publications that I was familiar with which contained analyses of ordinal data [e.g., 15, 17, 21, 34, 36] and conducted a search in the IEEE digital library using either the keyword *likert* or *ordinal regression* for papers published between the years 2025–2026 at IEEE VIS, the premier publication venue in the field of visualization (similar to the methodology used by South et al [27]). This search returned 60 results. However, as this yielded only two examples of studies which used ordinal regression models, I expanded the years to 2020–2026 only for the keyword *ordinal regression*. This resulted in a sample of 72 papers, of which 55 used a visualization to communicate the results of a Likert data analysis. I supplement the findings from my survey with the results of South et al.'s [27] systematic review of the use of Likert scale in Visualization papers. The qualitative analysis can be found in `supplement` ▶ survey. I chose this informal approach, as the objective of this survey was to get a sense of the diversity in analysis and reporting strategies that researchers have used and how researchers communicate the key takeaways from their Likert data, and not to quantify how frequently each modelling approach has been adopted.

3.1 When the Data is Modeled as Metric

Several of the papers in my survey (25/55) reported means and an uncertainty estimate (most commonly 95% confidence or credible intervals, but also in some cases standard deviations; see Figure 2A) for each item [e.g., 17, 26, 32]. Some researchers report the means and 95% intervals aggregated across multiple items [e.g., 4, 36], by taking the average for each participant, which is how Likert scales were original intended to be used [5, 14]. Other approaches that I observed were reporting the data distribution directly (Figure 2B) and comparing the means and the shapes of distribution. [e.g., 15], or using boxplots and densities [e.g., 18]. In South et al's review [27], 30/65 papers with a visualization communicate some form of mean and uncertainty estimate.

Researchers primarily used ordinal data to make comparisons between conditions by communicating mean point estimates and differences in means on the response scale. Even though a (hypothetical) statement such as "*participants rated visualization [X] to be more trustworthy (M = 4.1, SD = 1.05) compared to visualization [Y] (M = 3.6, SD = 1.4)*" may be technically flawed[5] when applied to data which is not on an interval scale as it entails arithmetic operations [31], it provides a sense of the data distribution and the average response in a particular condition are (at least approximately), and crucially how large the differences between conditions are *on the response scale*. For instance, Nadib et al. [17] reporting higher trust in one of their conditions contextualize this by adding "the effect, albeit statistically significant, is not large: trust increases by roughly half of a Likert point [...]."

3.2 When the Data is Modeled as Ordinal

I observed four distinct approaches for communicating the results of an ordinal model: (i) visualizing the probability of each ordinal category for each condition [12, 21, 24]; (ii) visualizing the empirical cumulative distribution function (CDF) plots [20]; (iii) showing the data distribution and calculating the odds ratio between conditions [2]; and (iv) estimating mean and 95% credible intervals from posterior draws (for a single item or averaged across multiple items) [7, 34]. The difference in how difficult it is to comprehend the results of ordinal models compared to metric models is stark—the first two approaches (Figure 2C-D), while more appropriate for the analysis, can make it significantly more difficult for a viewer to determine how much higher/better one condition is rated on average, compared to other conditions, in terms of points on the Likert scale (a straightforward task using the mean and interval plots).

This might be perhaps why Yang et al. [34] and Dragicevic et al. [7] adopted the approach of first modelling the data as ordinal, and then treating it as metric by estimating the mean of posterior

[4] A logit function can also be used as a link function in cumulative-link models. The role of the link function is to provide a one-to-one mapping between a continuous variable in $\mathbb{R}$ to $[0, 1]$. In theory, there are many such functions $f : \mathbb{R} \to [0, 1]$ that can be used as link functions. The choice of which link function to use is essentially arbitrary [35], and typically chosen based on convention and ease of interpretability. Here, I use the probit function due to mathematical convenience.

[5] Researchers also use statistical significance from ANOVAs to support their claims about which conditions are more highly rated, which is arguably more problematic—statistical significance tests should be conducted in conjunction with a power analysis, lest a researcher risks running overpowered studies, and many of the studies do. However, it is unclear whether the assumptions of power analyses for commonly used statistical tests hold for such skewed data distributions which are of concern here (e.g., ANOVA assumes that the data is normally distributed). Additionally, the ritualistic use of statistical significance tests as rituals [10, 11] is a broader issue which is out of scope of this current work.

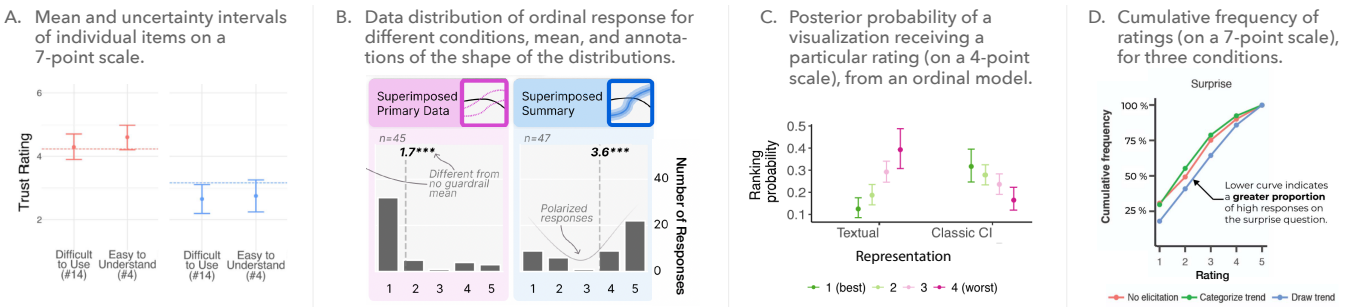

Figure 2: Prototypical examples of visualizations of ordinal data analyses from the survey. (A) shows mean and error bars representing standard errors (but other visualizations show 95% confidence or credible intervals) of individual items [32]. Similar visualizations were used to represent estimates aggregated across multiple items [e.g., 4, 34, 36]. (B) shows the complete data distribution with annotations of the mean and shape of the distribution [15]. (C) shows the posterior probability of a condition receiving a particular rating on a 4-point scale [12]. (D) Empirical cumulative distribution function of ratings (on a 7-point scale) for three treatment conditions [20]. Instead of using empirical frequencies, a similar visualization could also be created using posterior draws from an ordinal regression model.

predicted probability[6] across the ratings—it allowed them to communicate results on the response scale. While the approach by Rogha et al. [20] (see Figure 2D) visualizes the empirical cumulative distribution function (i.e., the probability that a participant rated at least $[k]$ on the 7-point Likert scale) for each condition on the response scale, it remains challenging and counter-intuitive to interpret—to determine which condition is rated higher, the viewer has to identify the curve which is *visually lower*, which runs counter to the typical convention in visualization of encoding higher as better. This unintuitiveness of the ordinal model has also led to an arguably idiosyncratic decision for reporting results in my own prior work [21]—I reported the probability of participants rating *3 or higher* on a 5-point Likert scale to compare across conditions, with the choice of 3 or higher being arbitrary (although hopefully justifiable). More conventionally, Bradley et al.'s [2] decision to report the odds ratio is similar to recommendations made by Syiem and Velloso [31]; they even augment this by converting the odd's ratio to a Cohen's d value. However, despite all this, the odds ratio between two variables on a latent (inverse probit) scale is significantly less intuitive than information communicated on the response scale.

Researchers also often communicated their key takeaways based on the proportion of responses that fell into, or above, a specific category, without modelling the data as either metric or ordinal, and using visualizations of the distribution of Likert responses such as histograms, stacked bar charts or heatmaps (26/55 papers in my survey and 27/65 papers in [27]). This takeaway, which is analogous to estimating $Pr(y_i \geq k)$, can be easily obtained using an ordinal model; however, this value can be quite difficult to estimate using either of the visualizations used in prior work (Figure 2C-D).

4 Towards a Better Visualization of Ordinal Regression Models

4.1 Design Properties

Based on the discussion above, I outline a set of design properties for visualizing the results of ordinal regression models which will make it intuitive for a viewer to easily extract the key pieces of information they care about:

[D1] The visualization should communicate the central tendency on the response scale (e.g., what is the model estimated response for the median participant?

[D2] The visualization should communicate uncertainty in the central tendency on the response scale (e.g., what is the uncertainty in the median estimate?

[6] In Dragicevic et al. [7], the posterior probability for each rating is multiplied by the "magnitude" of the rating, $\sum_k \mathbb{P}(y_i = k) \cdot k$, to obtain the posterior mean and credible interval

[D3] The visualization should easily allow the reader to identify which condition is rated "better." Consistent with most convention, I argue that the visual representation should encode conditions which are "better" higher (unlike the CDF plot in Figure 2D).

[D4] The visualization should allow comparison between conditions on the response scale (e.g., how many points higher does the median participant in one condition rate than the median participant in another condition?).

I propose the use of a *modified* Complementary Cumulative Distribution Function, which fulfils these design properties, as a better visualization of the results of ordinal models

4.2 The modified Complementary Cumulative Distribution Function (mCCDF) Plot

The Complementary Cumulative Distribution Function (CCDF) can be calculated from a cumulative probit-link regression model using a linear transformation: $\text{CCDF} = \Phi^{-1}(y_i > k) = 1 - \Phi^{-1}(y_i \leq k)$. Thus, the CCDF tells us the probability of an item being rated greater than category k. However, for Likert data, we care about the probability of an item being rated at category k or higher, which can be obtained using the following modified CCDF:

$$\text{mCCDF} = \Phi^{-1}(y_i \geq k) = 1 - \Phi^{-1}(y_i < k)$$

While seemingly trivial to estimate, the mCCDF has properties which make it more suitable to communicate the results of ordinal regression models, especially for single items. The rating corresponding to the intersection of the mCCDF draws with $y = 0.5$ represents the median rating for a particular condition. In Figure 3,

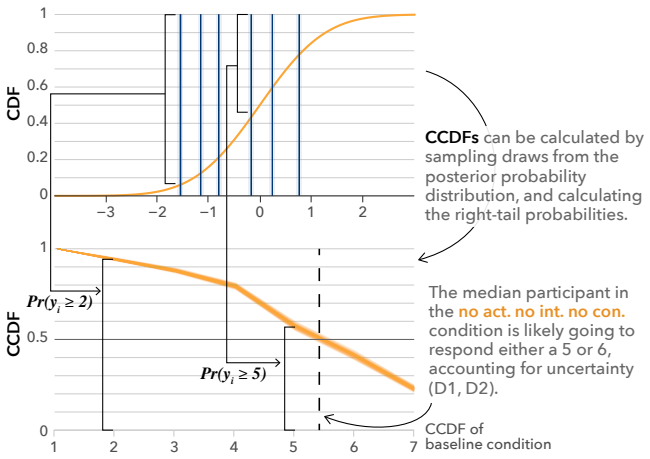

Figure 3: Calculation of a mCCDF plot from the latent variable and estimated thresholds.

the median rating is likely going to be a 5 or 6, even after taking uncertainty into account (D1, D2). The uncertainty in the model estimates is relatively low for this model and dataset due to the large number of responses collected.

In addition, the mCCDF more directly encodes one of the most relevant pieces of information for ordinal data—the probability that participants rated k or higher ($\mathbb{P}(y_i \geq k)$), unlike the conventional CDF plot which encodes $\mathbb{P}(y_i \leq k)$. This makes it more intuitive to perform two types of comparisons: (i) how much more likely are participants in one condition to rate k or higher relative to another (D3), as shown in Figure 4B2 (e.g., the probability of the *action, intention, no contact* condition being rated 6 or higher is 0.3 less than the baseline condition); and (ii) how much better one condition is relative to another (D4), based on the median participant (e.g., in Figure 4B(3), the median participant in the *action, intention, no contact* condition is going to rate at least a four, and the median participant in the baseline condition is going to rate at least a 5). Note that while these comparisons can be performed by a viewer using CDF plots as well, they are far less intuitive (for an approximation of the challenge, compare the magnitude and direction of the difference between conditions in Figure 4A).

An alternative approach for representing mCCDFs is visualizing them as step functions (a piecewise constant function). Representing mCCDFs as step functions has some trade-offs—reading the mean probabilities for each rating (and the associated uncertainty) for a particular condition becomes easier; however, determining the median rating for a condition, and thus comparing the median rating between conditions becomes more difficult. As such, the designer should choose which communication goal they wish to prioritise. [7]

4.3 Scales with Multiple Correlated Items

In some of the studies in the survey [4, 34, 36] responses from multiple, potentially correlated, items were aggregated and reported using a single point estimate and uncertainty interval. An analyst could instead fit a single hierarchical model to the data, with the correlated items modelled as random effects with different latent means and variances (referred to as *difficulty* and *discrimination* parameters in item-response theory). As scales with potentially correlated items are often measuring the same underlying construct, partial pooling of information in a hierarchical model allows parameters to be more efficiently estimated. We can then create mCCDF plots for each item individually, an "average" item (i.e., when the latent mean is zero), or marginalized over all of the items. I demonstrate how such an analysis can be implemented, and how the corresponding mCCDF plots can be created for each item in `supplement` ▶ RScript ▶ `03-unequal-variance-ccdf.Rmd`.

An additional benefit of adopting such models is that they can be useful for evaluating the items themselves, by assessing the degree of similarity in responses to the different items. Analogous approaches are used in item-response theory to evaluate questions, based on the coefficients for the difficulty and discrimination parameters and using item characteristic curves (ICC), to create validated set of test questions (see Section 5-7 and Figure 11 in [9]). For ordinal models, mCCDFs can complement the information presented using ICCs (which show the modeled probability of responding at or above each category as a function of the latent trait) by visualizing how similar participants' actual responses to each item actually are, in the response scale.

4.4 Should We All Be Using Ordinal Models?

While I personally prefer to use ordinal models for Likert data, I do not hold strong beliefs that this is objectively the best way of analysing ordinal data, as I am yet to find compelling evidence to suggest that the conclusions would be different regardless of the choice of models. Liddell and Kruschke [13] try to demonstrate

[7] I personally tend to prefer the line plot over the step function plot.

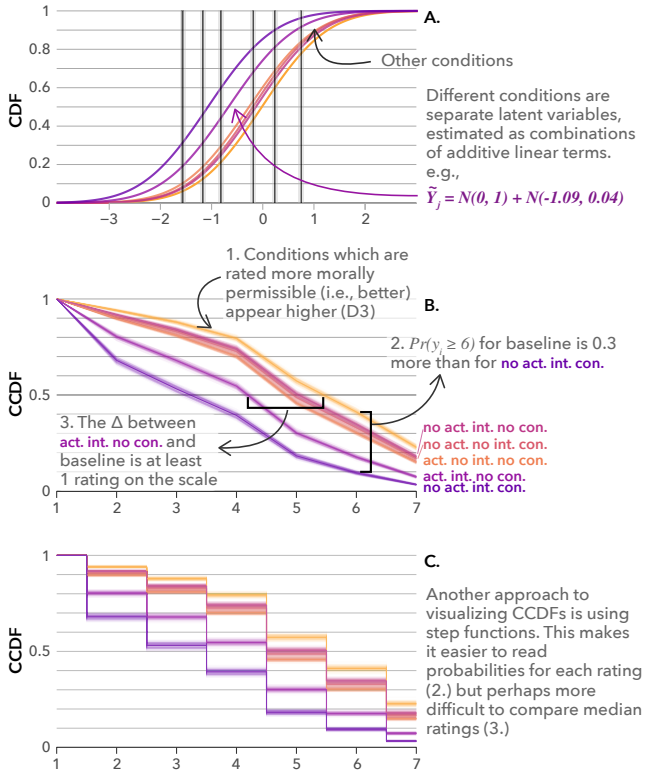

Figure 4: Calculation of a mCCDF plot for an experiment with multiple conditions (i.e., ordinal models with multiple predictors).

that ordinal regression is better, and that treating ordinal data as metric can lead to inflated false positive and false negatives, using simulated data. However, for their simulation, they generate ordinal data by discretizing a latent variable; thus, what they actually demonstrated was that the model which is consistent with their assumed data-generating process was better than a model which made obviously incorrect assumptions in the *small world* of their simulation.[8] In contrast, Dragicevic et al. [7] compare nine different data analysis approaches for the same dataset and found little difference in terms of takeaways between ordinal regression and metric-based approaches. My preference for using ordinal models is one based on statistical philosophy—I consider approaches which model the data directly, with as few assumptions as possible, to be better. The value of an ordinal model lies in the fact that it makes fewer assumptions regarding the data generating processes of Likert responses compared to metric-based approaches, which is why I would still recommend the use of ordinal regression model. However, the primary goal of this paper is not necessarily to urge researchers to adopt ordinal regression models but rather to demonstrate approaches for visualizing results of ordinal regression *if* researchers choose to adopt such models.

5 Conclusion

When analysing ordinal data, researchers appear to prefer to treat the data as metric and communicate the results—using point estimates, uncertainty, and differences between conditions—on the response scale. Complementary Cumulative Distribution Function (mCCDF) plots offer an approach for communicating the results of cumulative-link ordinal regression models on the response scale, allowing the reader to extract analogous pieces of information.

[8] The true data generating process is almost always unknown, and the models that we use to analyze data are *all* small world approximations [25], and thus rely on assumptions. The uncertainty regarding which model is correct has been described as ontological uncertainty [22, 23, 28].

Acknowledgments

I would like to thank Matthew Kay for detailed feedback on the visualization and the paper draft.

References

[1] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67:1–48, Oct. 2015. doi: 10.18637/jss.v067.i01 6

[2] D. Bradley, G. Strain, C. Jay, and A. J. Stewart. Magnitude Judgements are Influenced by the Relative Positions of Data Points Within Axis Limits. *IEEE Transactions on Visualization and Computer Graphics*, 31(2):1414–1421, Feb. 2025. doi: 10.1109/TVCG.2024.3364069 2, 3

[3] P.-C. Bürkner and M. Vuorre. Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, 2(1):77–101, Mar. 2019. doi: 10.1177/2515245918823199 2

[4] A.-F. Cabouat, T. He, P. Isenberg, and T. Isenberg. PREVis: Perceived Readability Evaluation for Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):1083–1093, Jan. 2025. doi: 10.1109/TVCG.2024.3456318 2, 3, 4

[5] J. Carifio and R. J. Perla. Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3(3):106–116, Mar. 2007. doi: 10.3844/jssp.2007.106.116 2

[6] F. Cushman, L. Young, and M. Hauser. The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm. *Psychological Science*, 17(12):1082–1089, Dec. 2006. doi: 10.1111/j.1467-9280.2006.01834.x 1, 6

[7] P. Dragicevic, Y. Jansen, A. Sarma, M. Kay, and F. Chevalier. Increasing the Transparency of Research Papers with Explorable Multiverse Analyses. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–15. ACM, Glasgow Scotland Uk, May 2019. doi: 10.1145/3290605.3300295 2, 3, 4

[8] J. Gabry, R. Češnovar, A. Johnson, and S. Bronder. Cmdstanr: R Interface to 'CmdStan', 2025. 6

[9] L. W. Ge, Y. Cui, and M. Kay. Calvi: Critical thinking assessment for literacy in visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581406 4

[10] G. Gigerenzer. Mindless statistics. *The Journal of Socio-Economics*, 33(5):587–606, Nov. 2004. doi: 10.1016/j.socec.2004.09.033 2

[11] G. Gigerenzer. Statistical Rituals: The Replication Delusion and How We Got There. *Advances in Methods and Practices in Psychological Science*, 1(2):198–218, June 2018. doi: 10.1177/2515245918771329 2

[12] J. Helske, S. Helske, M. Cooper, A. Ynnerman, and L. Besançon. Can Visualization Alleviate Dichotomous Thinking? Effects of Visual Representations on the Cliff Effect. *IEEE Transactions on Visualization and Computer Graphics*, 27(8):3397–3409, Aug. 2021. doi: 10.1109/TVCG.2021.3073466 2, 3

[13] 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, Nov. 2018. doi: 10.1016/j.jesp.2018.08.009 1, 4

[14] R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 22 140:55–55, 1932. 2

[15] M. Lisnic, Z. Cutler, M. Kogan, and A. Lex. Visualization Guardrails: Designing Interventions Against Cherry-Picking in Interactive Data Explorers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–19. ACM, Yokohama Japan, Apr. 2025. doi: 10.1145/3706598.3713385 2, 3

[16] R. McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. Chapman and Hall/CRC, New York, 2 ed., Mar. 2020. doi: 10.1201/9780429029608 1, 6

[17] K. A. Nadib, M. Kogan, A. Lex, and M. Lisnic. Guardrail Selection in Line Charts to Contextualize Persuasive Visualizations. *Computer Graphics Forum*, 2026. 2

[18] K. Nakano and T. Narumi. Avatars, Should We Look at Them Directly or Through a Mirror?: Effects of Avatar Display Method on Sense of Embodiment and Gaze. *IEEE Transactions on Visualization and Computer Graphics*, 31(5):2912–2922, May 2025. doi: 10.1109/TVCG.2025.3549545 2

[19] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, S. Heisterkamp, B. Van Willigen, and R. Maintainer. nlme : Linear and nonlinear mixed effects models. r package version 3.1-103. *http://cran.r-project.org/web/packages/nlme/index.html*, 2017. 6

[20] M. Rogha, S. Sah, A. Karduni, D. Markant, and W. Dou. The Impact of Elicitation and Contrasting Narratives on Engagement, Recall and Attitude Change With News Articles Containing Data Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 30(7):4375–4389, July 2024. doi: 10.1109/TVCG.2024.3355884 2, 3

[21] A. Sarma, S. Guo, J. Hoffswell, R. Rossi, F. Du, E. Koh, and M. Kay. Evaluating the Use of Uncertainty Visualisations for Imputations of Data Missing At Random in Scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):602–612, Jan. 2023. doi: 10.1109/TVCG.2022.3209348 2, 3

[22] A. Sarma, M. Hedayati, and M. Kay. More Forecasts, More (Decision) Problems: How Uncertainty Representations for Multiple Forecasts Impact Decision Making. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, pp. 1–14. Association for Computing Machinery, New York, NY, USA, Apr. 2025. doi: 10.1145/3706598.3713725 4

[23] A. Sarma, K. Hwang, J. Hullman, and M. Kay. Milliways: Taming multiverses through principled evaluation of data analysis paths. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642375 4

[24] A. Saske, L. Koesten, T. Möller, J. Staudner, and S. Kritzinger. A Multidimensional Assessment Method for Visualization Understanding (MdamV). *IEEE Transactions on Visualization and Computer Graphics*, 32(3):2695–2708, Mar. 2026. doi: 10.1109/TVCG.2026.3653265 2

[25] L. J. Savage. *The Foundations of Statistics*. Dover Publications, New York, 2d rev. ed ed., 1972. 4

[26] H. Song, A. Cho, C. X. Bearfield, and J. Stasko. Visualizing Trust: How Chart Embellishments Influence Perceptions of Credibility. *IEEE Transactions on Visualization and Computer Graphics*, 32(1):1306–1316, Jan. 2026. doi: 10.1109/TVCG.2025.3634785 2

[27] L. South, D. Saffo, O. Vitek, C. Dunne, and M. A. Borkin. Effective use of likert scales in visualization evaluations: A systematic review. *Computer Graphics Forum*, 41(3):43–55, 2022. doi: 10.1111/cgf.14521 2, 3

[28] D. Spiegelhalter. Risk and Uncertainty Communication. *Annual Review of Statistics and Its Application*, 4(Volume 4, 2017):31–60, Mar. 2017. doi: 10.1146/annurev-statistics-010814-020148 4

[29] S. S. Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684):677–680, June 1946. doi: 10.1126/science.103.2684.677 1

[30] R. C. Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2024. 6

[31] B. Victor Syiem and E. Velloso. Better Assumptions, Stronger Conclusions: The Case for Ordinal Regression in HCI. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pp. 1–21. ACM, Barcelona Spain, Apr. 2026. doi: 10.1145/3772318.3790821 1, 2, 3

[32] H. W. Wang, K. Lin, A. Cohen, R. Kennedy, Z. Zwald, C. Nobre, and C. X. Bearfield. Do You "Trust" This Visualization? An Inventory to Measure Trust in Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 32(3):2515–2528, Mar. 2026. doi: 10.1109/TVCG.2025.3646847 2, 3

[33] G. Wilkinson and C. Rogers. Symbolic description of factorial models for analysis of variance. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 22(3):392–399, 1973. doi: 10.2307/2346786 6

[34] F. Yang, M. Cai, C. Mortenson, H. Fakhari, A. D. Lokmanoglu, J. Hullman, S. Franconeri, N. Diakopoulos, E. C. Nisbet, and M. Kay.

Swaying the Public? Impacts of Election Forecast Visualizations on Emotion, Trust, and Intention in the 2022 U.S. Midterms. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):23–33, Jan. 2024. doi: 10.1109/TVCG.2023.3327356 2, 3, 4

[35] F. Yang, M. Hedayati, and M. Kay. Subjective Probability Correction for Uncertainty Representations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–17. Association for Computing Machinery, New York, NY, USA, Apr. 2023. doi: 10.1145/3544548.3580998 2

[36] F. Yang, C. R. Mortenson, E. Nisbet, N. Diakopoulos, and M. Kay. In Dice We Trust: Uncertainty Displays for Maintaining Trust in Election Forecasts Over Time. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–24. ACM, Honolulu HI USA, May 2024. doi: 10.1145/3613904.3642371 2, 3, 4

A Supplementary Materials

All supplemental materials are available on OSF at `https://osf.io/cdgby/`, released under a CC BY 4.0 license. In particular, they include (1) Excel files containing the data for the literature survey, (2) full R scripts for recreating the analysis and figures in the paper, (3) figure images, and (4) a full version of this paper with all appendices.

B Details on Dataset and Ordinal Model Used

B.1 Dataset

The dataset used in the analysis described in this paper was collected by Cushman et al. [6] in a series of experiments to collect empirical evidence on how people judge the "moral goodness or badness" of actions [16]. The experiments present participants with a scenario—variations of the trolley problem—and asks participants to rate the action (or inaction) on a scale from 1 to 7. The following is the traditional version of the trolley problem [16]:

> "Standing by the railroad tracks, Dennis sees an empty, out-of-control trolley about to hit five people. Next to Dennis is a lever that can be pulled, sending the trolley down a side track and away from the five people. But pulling the lever will also lower the railing on a footbridge spanning the side track, causing one person to fall off the footbridge and onto the side track, where he will be hit by the trolley. If Dennis pulls the lever the trolley will switch tracks and not hit the five people, and the one person to fall and be hit by the trolley. If Dennis does not pull the lever the trolley will continue down the tracks and hit five people, and the one person will remain safe above the side track."

The experiments by Cushman et al. [6] investigate the role of three principles from moral psychology and philosophy on moral judgements. The principles are:

1. **The action principle**: Harm caused by action is morally worse than equivalent harm caused by omission.
2. **The contact principle**: Using physical contact to cause harm to a victim is morally worse than causing equivalent harm to a victim without using physical contact.
3. **The intention principle**: Harm intended as the means to a goal is morally worse than equivalent harm foreseen as the side effect of a goal.

The above scenario applies the *action* principle as the actor has to perform an action to create the outcome (as opposed to being passive and letting events unfold). However, the other two principles do not apply as there is no contact (i.e., no *contact*), and the harm caused to the person on the footbridge is accidental and not intentional (i.e., no *intention*). The three principles can be varied to create different scenarios. In the study by Cushman et al. [6], participants "received 32 moral scenarios separated into two blocks of 16. Each block included 15 test scenarios and 1 control scenario." The dataset contains responses from 331 participants, and includes details on participants' age, gender, and level of education. For more details on the dataset, please refer to [6, 16].

B.2 Analysis

The mCCDF visualizations described in this paper are based on the results of the analysis outlined by McElreath [16], which applies a Bayesian ordinal cumulative-logit model. However, I instead use the probit link function (as opposed to the logit link function used by McElreath [16]), due to mathematical convenience. The model can be described using the Wilkinson-Rogers-Pinheiro-Bates syntax [1, 19, 33] as:

probability of data

`line 1` $y \sim \text{Categorical}(p_i)$

probabilities of each value of k

`line 2` $p_1 = q_1$

`line 3` $p_j = q_j - q_{j-1} \quad \text{for } 1 < j < K$

`line 4` $p_K = 1 - q_{K-1}$

`line 5` $\text{probit}(q_j) = \tau_j - \phi_i$ — cumulative probit link

`line 6` $\phi_i = \beta_{A[i]} + \beta_{C[i]} + \beta_{I[i]} + \beta_{A[i]} \cdot \beta_{I[i]} + \beta_{C[i]} \cdot \beta_{I[i]}$ — linear model

`line 7` $\beta \sim \text{Normal}(0,1)$ — prior for coefficients

`line 8` $\tau_j \sim \text{Normal}(0,1)$ — prior for cutpoints

`line 9` $i \in \{1...N\}$ (N *participants*)

`line 10` $k \in \{1...K\}$ (K number of ordered categories)

Line 1: The rating the participants assigned is modeled as a Ordered categorical distribution.

Line 2-4: The categorical distribution takes a vector of probabilities $p = p_1, p_2, p_3, p_4, p_5, p_6$ of probabilities of each response value below the maximum response (p_7).

Line 5: Each response value k in this vector is defined by its link, using the probit function to an intercept parameter, τ_k and the linear model ϕ_i.

Line 6: For each response, we assume that participants' ratings depend on the condition that the principle (or the combination of principles) the scenario is based on.

Priors: We use weakly-regularizing priors centered on zero (no effect) while permitting the possibility of significant distortion: $\beta, \tau_j \sim \text{Normal}(0,1)$.

Implementation: We implemented these models in `R 4.4.0` [30] and `CmdStanR 0.8.0` [8]. The model ran for four chains with 5,000 warmup samples and 5,000 post-warmup samples each, thinned by 4 for a final total sample size of 5,000. We assessed convergence using the Gelman-Rubin diagnostic ($\hat{R} = 1.00$ for all population-level parameters, correlations and standard deviations) and the (bulk and tail) effective sample sizes ($\text{ESS}_{min} \approx 2,000$).